Rehnüma Character Analysis

Design, Validation Protocol and Everyday-Life Dashboard Integration of a Personality and Vocational Orientation Test Built on Open-Source Psychometric Instruments

Technical monograph · Version 1.2 · 22 November 2026

Rehnüma (working name: LifeOS) · rehnuma.co

Abstract

This monograph documents the design and validation protocol of the pre-sign-up "character analysis" test of Rehnüma, a personal operating system, and the method by which the test results are carried into the product's everyday-life widgets. The test is built on three published, openly accessible instruments: the IPIP-NEO-120 (Five-Factor Model, 30 facets), the O*NET Interest Profiler Short Form (Holland RIASEC) and the O*NET Work Importance Locator (work values), plus a three-question rhythm (chronotype) self-report. The instruments were adapted into Turkish, arranged in a two-stage form (60 + 60 items on the personality scale) and equipped with response-quality indicators. The current form (v5) adds two ranking steps without changing the scales: self-placement on eight self-type cards and an eight-round statement ranking (ADR-044). Results are reported both as continuous scores and as derived narrative layers (mind motif, temperament, eight self types, natural tendencies); each derived label is presented with an uncertainty marker. The report has fourteen sections and is delivered in five languages (Turkish, English, French, German, Arabic); personality percentiles are based on the published international sample of the IPIP-NEO-120 (Johnson, 2014; 619,150 respondents; ADR-082).

Validation is tied to a five-gate protocol. At Gate 0, 300 synthetic respondents completed both stages of the test with form v4 through the actual service, and 60 of them retook it: facet recovery median r = .89, domain r = .87–.89, RIASEC r = .92–.94; careless-responder detection sensitivity 1.00. The same simulation quantified the fragility of categorical labels: with the true trait held constant, the four-letter code remained the same in 57% of cases, the first Holland letter in 45% and the top archetype in 70%. These findings were written back into the design by raising the "balanced" threshold to 8 points, showing alternative motifs and adding the "neck-and-neck" note. Psychometric evidence for the Turkish version (Gates 1–4) will be collected from real user data; report language does not carry the claim of a gate that has not been passed.

Chapter 6 examines with what evidence personality findings can be linked to 229 decision scenarios across seventeen product surfaces. A six-strand, source-cited literature review restricted the variables that may serve as rule conditions to Emotional stability, Conscientiousness, rhythm and interests, leaving mind motif, temperament, archetype and work values at the narrative level. The outcome is a machine-readable rule catalogue of 341 records and a policy with three statuses assigned by evidence

grade (rule, A/B candidate, no change); personality-conditioned rules only soften language or add a suggestion text to an otherwise empty slot, and never hide information. Chapter 7 documents the Five-Branch Character Matrix, a corpus of 76 cells backed by 209 sources that, under the same evidence policy, links personality domains and facets to behaviour in five life branches (finance, health, relationships, career, spiritual life) (ADR-037).

Keywords: Five-Factor Model, IPIP-NEO-120, Holland RIASEC, O*NET, test adaptation, simulation, Barnum effect, personalisation, chronotype, evidence-based design.

Abbreviations

Abbreviation	Expansion
A/B	Comparative trial of two versions
ADR	Architecture Decision Record (decision record, `docs/decisions/`)
BFI-2	Big Five Inventory-2
CFA	Confirmatory Factor Analysis
DIF	Differential Item Functioning
EFA	Exploratory Factor Analysis
EFPA	European Federation of Psychologists' Associations
HEXACO	Honesty-Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, Openness
IPIP	International Personality Item Pool
IP-SF	O*NET Interest Profiler Short Form
ITC	International Test Commission
KVKK	Kişisel Verilerin Korunması Kanunu (Turkish Personal Data Protection Law No. 6698)
MBTI	Myers-Briggs Type Indicator
MCTQ	Munich ChronoType Questionnaire
MEQ / rMEQ	Morningness-Eveningness Questionnaire / its reduced form
O*NET	Occupational Information Network (USDOL/ETA)
RCT	Randomised controlled trial
RIASEC	Realistic, Investigative, Artistic, Social, Enterprising, Conventional
SDS	Self-Directed Search (Holland)
TRAPD	Translation, Review, Adjudication, Pretesting, Documentation
USDOL/ETA	U.S. Department of Labor, Employment and Training Administration
VIA	Values in Action (classification of character strengths)
WIL	O*NET Work Importance Locator

Contents

1. Introduction
2. Theoretical Framework and Literature

3. Measurement Design
4. Validation Protocol and Simulation Findings
5. Reporting the Results
6. Character Trace: Carrying Personality Findings into the Everyday-Life Dashboard
7. The Five-Branch Character Matrix: Carrying Personality into Behaviour across Life Domains
8. Ethics, Privacy and Licences
9. Discussion
0. Conclusion and Future Work

References · Appendices

Version and source

This document is compiled from the Markdown chapters under `docs/research/tez/` (`tooling/build.mjs`). The decision records, research notes and source code on which each chapter rests are in the same repository; file paths in the text are relative to the repository root. Version 1.0 (23 October 2026) documents the pre-pilot state; Version 1.1 (25 October 2026) adds the chapter on the Five-Branch Character Matrix (Chapter 7). Version 1.2 (22 November 2026) brings the text into line with the state of the system on that date: form v5, percentiles from the published norm, the fourteen-section report and the five languages have been incorporated, and chapter numbers and figures have been synchronised with the code. A new version is published as validation gates are passed, and earlier versions are preserved. This English edition is a translation of the Turkish Version 1.2; where the two differ, the Turkish text is authoritative. It is compiled from `docs/research/tez/en/` with `node build.mjs --lang en`.

1. Introduction

This work addresses the design of a personality and vocational orientation test developed for the personal operating system named Rehnüma, its validation protocol, and the carrying of the findings the test produces into everyday-use surfaces. Rehnüma is a web application that brings goal (OKR), habit, health and finance modules together under a single data model and a single experience. The test is positioned as the pre-sign-up value of this product: the visitor takes the test before seeing the dashboard; the account step is at the end of the test, scoring is done after the account has been created, and the result is shown on the dashboard (ADR-025; ADR-073). The scoring, occupation matching and guidance core of the test runs as a separate service named `Pusula-API` (ADR-027); the form has a two-stage structure (ADR-026; current form v5, ADR-044), the depth of the report is derived from measurement at facet level (ADR-028), and the psychometric validation of the Turkish version is carried out in stages under a five-gate protocol (ADR-029).

The thesis of the work is, briefly, this: the "type" certainty offered by most popular tests on the market cannot be defended scientifically, whereas a test that is built on published open-source instruments tested in large samples, that states its uncertainty openly and that links its findings to everyday decisions according to evidence grade is both more useful to the user and more distinctive than its competitors. This chapter defines the problem, sets out the research questions and contributions, delimits the scope and explains the organisation of the work.

1.1 Problem

1.1.1 The market for pre-sign-up personality tests

Consumer-facing personality and career tests largely follow the same funnel pattern: the test is taken first, a short free result is shown, and then a paid full report is offered. At Truity no sign-up is required for the test and the short result, and the account is created after the test (Truity, n.d.). 16Personalities frames the free result as "about 5% of what we can show" and sells the full reports as separate packages (16Personalities, n.d.-b). CliftonStrengths, by contrast, offers no free tier; payment and an access code come before the test (Gallup, n.d.). In Turkey, products aimed at students are sold at between 289 TL and 990 TL, and a product that adds a mapping to ÖSYM programmes for university choice is sold at 3,500 TL (Kariyer Planım, n.d.; Remetesting, n.d.; CharacterIX, n.d.). This price range indicates that the user expects from a test information that bears on life decisions (inference).

A pre-sign-up test is attractive as the first value of a product; but this position creates two pressures. The first is the pressure of duration. In web surveys most of the dropout occurs at the beginning; thereafter the increase per item is moderate (Hoerger, 2010). The duration that respondents regard as ideal is 10–15 minutes, the maximum they regard as acceptable is 20–25 minutes (Revilla & Höhne, 2020), and abandonment on mobile devices is markedly higher than on desktop (Mavletova & Couper, 2014). The second is the pressure to persuade: the wish to produce a result that looks "rich" enough to trigger the decision to sign up and pay. This pressure easily turns into producing labels and template texts of a certainty that the measurement cannot bear. One of the most frequent complaints

in user reviews of competing products is that the paid report offers generic content not personalised to the person's responses (Trustpilot, n.d.); another complaint is encountering an unexpected paywall at the end of a promise of something free.

The direct examination that the authors carried out on the 16Personalities test on 23 October 2026 provided a concrete example of this pressure: the text shown in the locked sections of the paid full report was a filler text with its letters scrambled; that is, what was presented to the user in blurred form was not the real content behind the lock but a fake text giving the impression that content existed (details in Chapter 2.11). This is a striking observation on where pre-sign-up tests stand on the line between "showing value" and "making it look as though there is value".

1.1.2 Popular tests with a weak scientific basis, and type models

A substantial share of the most visible tests on the market rest on type models: the four-letter codes that come from the tradition of the Myers-Briggs Type Indicator (MBTI), Keirsey's four temperaments and the nine types of the Enneagram. The common problem of these models is that they divide continuously distributed personality traits into categories. McCrae and Costa (1989) found no support for the view that the MBTI really measures dichotomous preferences or qualitatively distinct types; they showed that the four indices measure four of the five dimensions of the Five-Factor Model. This finding was replicated in a sample of 900 people (Furnham et al., 2003). Retest instability at type level is also high: it is reported that, over a five-week interval, about half of the participants were classified differently on at least one dimension (Pittenger, 2005). For the Enneagram, a systematic review covering 104 samples found that factor analyses generally find fewer than nine factors and that there is no study deriving the nine types by clustering (Hook et al., 2021).

The problem is not only theoretical; it is also visible in the product's own measurement. In the simulation run on the Rehnüma test with form v4, with 300 synthetic respondents and 60 retests, the four-letter code derived from the Big Five scores remained the same in only 57% of cases even though the latent trait did not change at all between the two administrations (ADR-029). The variability came not from the scale but from the cut-off point: letters far from the boundary never flipped, while 36% of the letters close to the boundary flipped. This result is the product-level reflection of the general methodological finding that splitting a continuous variable in two loses information about individual differences (MacCallum et al., 2002).

Despite these weaknesses, type models answer a product need: a memorable, shareable identity. This need of the user is legitimate, and the question of the work is not to ignore it but to meet it with the honesty that the measurement can bear. Moreover, a rich report increases the risk of the Barnum effect: people accept with high rated accuracy statements that are general, positive and presented as "specific to you" (Dickson & Kelly, 1985), and the perceived accuracy of a real personality summary may not be distinguishable from that of general and socially desirable summaries (Poškus et al., 2014). Consequently, the datum "the user found the report accurate" cannot by itself be counted as evidence of the validity of the report.

1.1.3 The lack of open-source adapted instruments in Turkish

Instruments with a strong scientific basis exist, but those that can be used in a commercial Turkish web product are limited. Among the widely used Big Five scales that measure at facet level, the BFI-2 is open without permission only for non-commercial research and educational use (Soto & John, 2017); the site of the HEXACO inventory prohibits online administrations open to the general public even for researchers (HEXACO, n.d.). When licence and facet coverage are considered together, what remains is the IPIP-NEO-120, derived from the public-domain IPIP pool (Johnson, 2014; IPIP, n.d.-a). On the interest side, the short forms of the O*NET Interest Profiler are offered under a royalty-free developer licence; but translation and shortening make a validation study before the product goes live, and a mandatory notice of modification to the user, a condition (O*NET Resource Center, n.d.-b).

On the Turkish side, no published adaptation of any of these instruments could be found. The only Turkish record on the IPIP site is a translation of a 924-item pool, and the IPIP project has not validated this translation (IPIP, n.d.-b); the Beş Faktör Kişilik Envanteri (Five Factor Personality Inventory) derived from this pool does not follow the IPIP-NEO structure (Somer et al., 2002). For RIASEC, the Turkish scales either do not correspond one-to-one with Holland's six areas or, by their authors' own assessment, require substantial revision (Süerdem & Erkök, 2016; Perkmen & Tezci, 2015). No publicly available Turkish RIASEC norm broken down by age group could be found. None of the consumer-facing test products in Turkey publishes a reliability or validity figure; among global consumer products only two examples that give coefficients on their pages were identified (16Personalities, n.d.-c; Traitify, n.d.).

This gap has two consequences. First, to say that a Turkish test is "scientific" means transferring the English evidence of the source instrument to the translation, and this transfer is not automatically valid; translation can shift the content of a facet into a neighbouring domain, and in adolescents acquiescence bias can distort the scale structure (Soto et al., 2008). Second, conducting an honest validation process in the local market and publishing the figures is a differentiator in its own right (inference).

1.1.4 The failure to carry the result into everyday life

The report of a personality test is generally read and put aside. In none of the competing products reviewed was the report seen to be linked to daily tracking, to goals or to a habit plan. Yet adding guidance on how weaknesses may be overcome to written personality feedback increases the intention to use the feedback (University of Connecticut, n.d.), and in the meta-analysis of career interventions individualised interpretation was found to be one of the critical ingredients contributing to effectiveness (Brown & Ryan Krane, 2000).

In the context of Rehnüma this transfer is a natural question: given that goal, habit, sleep, training and budget surfaces already exist for the same user, can the findings of the test change the behaviour of these surfaces? ADR-025 opened this idea with the examples "a smaller step at low Conscientiousness, the time of day according to chronotype, gentle language at low Emotional stability". However, the evidence base for the transfer is far thinner than the literature on the relation of personality to behaviour. It has been shown meta-analytically that the Big Five domains predict goal-setting and self-efficacy motivation (Judge & Ilies, 2002); by contrast, no randomised study could be found that tests, with a behavioural outcome, whether a message tailored to personality changes behaviour more than an untailored message (Character trace evidence map report, `docs/research/re-`

`ports/`). Personalisation that goes beyond the limit of the evidence carries measurement error into product behaviour and makes the user live out the consequence of a label. The problem, therefore, is to determine where the link should not be made as much as to make the link.

1.2 Aim and research questions

The aim of the work is to set out the design principles, the validation path and the reporting rules of a personality and vocational orientation test that is offered before sign-up, rests on open-source instruments, states its uncertainty openly and links its findings to the decisions of a life dashboard according to evidence grade. This aim divides into four research questions.

RQ1. With which open-source instruments can an ambitious yet honest test be built? "Ambitious" here means not more type layers but measurement resolution: thirty facets alongside the five domains, the full profile of the six interest areas, work values and chronotype. "Honest" means that the licence of each layer is suitable for commercial use, that the reliability in its source is known and that the report does not go beyond this reliability. The question covers scale selection, licence constraints, the time budget and the two-stage form structure.

RQ2. How are validity and reliability demonstrated in stages? The English evidence of the source instruments does not pass automatically to the Turkish version. What claim can a product make before user data have accumulated; which new claim does each additional data threshold open? The question requires thinking, in sequence, about the translation and adaptation method, the verification of scoring by simulation, pilot item analysis, construct validity, retest stability and the construction of norms.

RQ3. How is the result reported without falling into the Barnum effect? The user expects the depth seen in competitors (type names, strengths, areas for development, suggestions for working together); this depth is easily imitated with general and positive sentences. The question covers tying every sentence to a scale and a score band, communicating measurement error and label instability to the user, and measuring distinctiveness rather than perceived accuracy.

RQ4. With what evidence can personality findings be linked to the widget decisions of a life dashboard? Which of the variables the test produces can change which widget behaviour, and with what evidence grade; where should nothing change; how is this link kept in code in an auditable form? The question asks about the limits of translating the behavioural literature of personality psychology into product rules.

Table 1.1. Research questions, the chapters in which they are answered and the main types of evidence.

Question	Focus	Main chapters	Type of evidence
RQ1	Instrument selection, licence, form structure	2, 3, 8	Published psychometric reports, licence texts, the abandonment literature
RQ2	Staged validation	4	International test adaptation guidelines, simulation (Gate 0), pilot plan
RQ3	A Barnum-resistant report	5	The feedback and Barnum literature, ITC and EFPA report criteria
RQ4	Transfer to the dashboard	6, 7	The personality × behaviour literature, the inventory of 17 surfaces, the rule catalogue, the five-branch character matrix

Note. Chapter numbers follow the organisation in Table 1.2. The ethics, privacy and licence dimension is addressed separately in Chapter 8 as a constraint cutting across every question.

1.3 Contributions

The contributions of the work come from a product study that rests on the literature like a thesis but whose source code and decision documents can be inspected. They can be gathered under six headings.

First, a measurement architecture built by reading licence and psychometrics together. We selected the IPIP-NEO-120 for the Big Five, the O*NET Mini-IP and Short-IP for interests, a card ranking in the tradition of the O*NET Work Importance Locator for work values and, for chronotype, short preference questions that are not presented as a psychometric scale until licence permission arrives. The selection was justified together with the exclusion of scales that are closed to commercial use or whose licence is unclear (BFI-2, HEXACO, VIA, Grit-S, SWLS) (Chapter 2.3). This architecture was placed in a two-stage form: a core of about 12–13 minutes before sign-up yields the domain scores reliably, and the deepening after sign-up, without asking the same items again, completes the facets to four items each and the interest scale to 60 items (ADR-026). The free summary and the full report thus become two resolutions of the same measurement. Form v5 adds two ranking steps without touching these scales: self-placement on eight self-type cards at the start of the first stage, and an eight-round statement ranking at the start of the second stage (ADR-044).

Second, a five-gate validation protocol and a simulation run before any user data. The protocol ties to each gate a condition, a numerical threshold and the claim that opens when that threshold is passed (ADR-029). At Gate 0 we tested the scoring code on the actual `Pusula-API` with 300 synthetic respondents and 60 retests; by measuring the structural fragility of the derived labels, the same simulation made the distinction between "the translation is broken" and "the label is fragile anyway" possible before the real pilot (`docs/research/scripts/simulate.py`). This is an example of a product knowing the limit of its own measurement numerically before it reaches the user.

Third, giving type models an honest place within the product. The four-letter code, the temperament and the archetypes were defined not as the main output but as secondary narrative layers that are derived from measured continuous scores and that state the clarity of each letter, the "balanced" state and the possibility of change on retest (ADR-026; ADR-028). The Enneagram was excluded entirely. The critique of types in the literature was combined with the values of label instability measured by simulation.

Fourth, a practicable reporting arrangement against the Barnum effect. Every report sentence is tied to a scale, a facet and a score band; an opposite sentence exists for the opposite band; personal lines are written only where there is a marked deviation in the facet; and in every section statement-level "Does this fit you?" feedback is collected so that non-distinctive sentences can be weeded out (ADR-028; Chapter 5).

Fifth, an evidence policy and a rule engine for carrying personality findings into product behaviour. The 229 scenarios defined across 17 widgets and surfaces were matched with a six-strand evidence map study and turned into a rule catalogue of 341 records (ADR-033). The essence of the policy is this: the condition of open rules today is only Emotional stability (and its facets),

Conscientiousness, chronotype, the first interest area and response quality; records tied to the other domains and facets are closed A/B candidates; mind motif, temperament, archetype, tendency and work values cannot trigger any widget rule. Personality-conditioned rules may only reduce or soften; those that propose a different value remain closed A/B candidates. This prohibition is kept as an executable invariant in `Back-end/tests/Feature/InvariantsTest.php`.

Sixth, a consolidated literature and licence map for the Turkish context. Turkish adaptations of the Big Five, RIASEC and chronotype; the commercial-use provisions of the IPIP, O*NET, BFI-2, HEXACO, VIA and other scales; and the price, flow and evidence status of test products in Turkey and on the global market were compiled in a single framework (Chapter 2).

None of these contributions includes the claim that the Turkish version has been psychometrically demonstrated. As of the date stamp of the work the correct statement is this: the test is built on published open-source instruments tested in large samples; the psychometric validation of the Turkish version is continuing with the pilot (ADR-029: Gate 0 passed, Gates 1–4 pending).

1.4 Scope and exclusions

The work covers the personality and vocational orientation test within the Rehnüma web product; the scoring, profile and occupation matching logic of the test in the `Pusula-API` service; the report and rule engine layer in the Rehnüma back end; and the test and report interface in the front end. The target audience is upper-secondary and university students and working or job-seeking adults, distinguished by a status question; a parental consent flow is envisaged for users under 18 (ADR-025).

Three areas have been deliberately excluded.

The mobile application. The Flutter-based mobile application of the Pusula product, from which the test originates, is not part of this study. The narrative texts and screen order of the mobile screens were taken only as a specification, and the web interface was rewritten in React (ADR-025). The form v4 improvements do not pass automatically to the mobile application (ADR-027).

The live Pusula service. In the first design the test was run through a bridge that connected to the live Pusula back end with shadow accounts. This route was abandoned because it put the mobile application at risk and created unnecessary identity complexity; the test, scoring and occupation matching core was forked within the repository as `Pusula-API` (ADR-027). The code, data and users of the live Pusula are outside the scope of the work.

Payment. While the test was being designed, whether the full report would be paid for or included in the membership was an open product question, and the place of payment in the flow had been left empty (ADR-025). The sales model was determined later in separate decision documents: the analysis gives the open sections of the report, all sections are unlocked with the annual dashboard membership and the full report is not sold separately (ADR-041; ADR-083). Pricing, distance selling and refund conditions are not the subject of this work; market prices are addressed only in the context of the competitor analysis, and the ethical side of the sales model in Chapter 8.

In addition to these, the following topics are also out of scope: use of the test for recruitment or selection (this use gives rise to a separate validation obligation under the O*NET licence and is contrary to the product's definition of low-stakes use in which the person obtains information for his or her own

interest); clinical diagnosis or mental health screening; measurement systems such as the Enneagram and VIA that are excluded on scientific or licence grounds; and a logging and entry flow on Rehnüma's health page (the health page shows the plan; tracking is in the habits module). The legal assessment concerning the protection of personal data is discussed as a framework in Chapter 8; it is not a substitute for legal advice. The interface and the report are offered in five languages (ADR-046); the validation discussion of the work is limited to the Turkish version, and the French, German and Arabic item translations have not been validated.

1.5 Organisation of the work

The work consists of ten chapters and appendices. The chapters follow the order of the research questions: first the theoretical ground and the measurement design (RQ1), then validation (RQ2), reporting (RQ3) and, in two chapters, transfer to the dashboard (RQ4); the chapter on ethics, privacy and licences gathers the constraints common to these four questions.

Table 1.2. The chapters of the work and their contents.

Chapter	Title	Content
1	Introduction	Problem, aim, research questions, contributions, scope
2	Theoretical Framework and Literature	The Five-Factor Model and the IPIP-NEO; alternative scales and their licences; Turkish adaptations; RIASEC and O*NET; work values; chronotype; type models and their critique; response quality; abandonment and duration; products on the market
3	Measurement Design	The two stages of the form (v4) and the ranking steps of v5, item pools, translation and adaptation, response scale, attention-check items, keying balance, scoring, derived layers, occupation matching, percentiles and the published norm
4	Validation	Validity and reliability framework, the five-gate protocol, the Gate 0 simulation and its findings, decisions taken on the basis of the findings, the analysis plan for Gates 1–4
5	Reporting	The distinction between summary and full report, the Barnum check, communicating uncertainty, percentile honesty, the information architecture and visual language of the report, trust framing
6	Character Trace	Inventory method (17 surfaces, 229 scenarios), evidence map, evidence policy, rule catalogue and engine, a worked sample profile
7	The Five-Branch Character Matrix	The two-layer principle, the literature review in five fields, corpus and contract, transfer to the dashboard and the report, limits
8	Ethics, Privacy and Licences	Consent, under-18s, data minimisation, 30-day deletion, service separation, licence obligations, health and finance limits
9	Discussion	Answers to the research questions, relation of the findings to the literature, limitations, tensions between product and science
10	Conclusion and Future Work	Conclusion, future work

Note. The consolidated reference list and the appendices (index of decision records, research notes, reproduction commands, sample rule catalogue record, public documents) are at the end of the work.

Two notes should be made for the reader. First, the work is a product study and most of the decisions rest on the decision documents (ADRs) in the repository; these documents are cited in the text in the form "(ADR-029)", and code is cited by file path. The reader can trace every claim down to the source code and the decision record. Second, the work deliberately restricts its language of evidence: every

sourced finding is marked with a citation, the authors' interpretation is marked as "inference", and the evidence grade is stated as strong, moderate, weak or none. Claims belonging to validation gates not yet passed for the Turkish version do not enter the text.

Chapter references

16Personalities. (n.d.-b). *Premium*. https://www.16personalities.com/premium

16Personalities. (n.d.-c). *Reliability and validity*. https://www.16personalities.com/articles/reliability-and-validity

Brown, S. D., & Ryan Krane, N. E. (2000). Four (or five) sessions and a cloud of dust: Old assumptions and new observations about career counseling. In S. D. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed.). Wiley. https://ncda.org/aws/NCDA/pt/sd/news_article/5310/_PARENT/CC_layout_details/false (reference to be verified)

CharacterIX. (n.d.). *Mesleki tercih*. https://characterix.global/CareerChoice/TR

Dickson, D. H., & Kelly, I. W. (1985). The "Barnum effect" in personality assessment: A review of the literature. *Psychological Reports, 57*(2), 367–382. https://doi.org/10.2466/pr0.1985.57.2.367

Furnham, A., Moutafi, J., & Crump, J. (2003). The relationship between the revised NEO-Personality Inventory and the Myers-Briggs Type Indicator. *Social Behavior and Personality, 31*(6), 577–584. https://doi.org/10.2224/sbp.2003.31.6.577

Gallup. (n.d.). *CliftonStrengths (StrengthsFinder)*. https://www.gallup.com/cliftonstrengths/en/strengthsfinder.aspx

HEXACO. (n.d.). *The HEXACO Personality Inventory – Revised*. https://hexaco.org/hexaco-inventory

Hoerger, M. (2010). Participant dropout as a function of survey length in Internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior, and Social Networking, 13*(6), 697–700. https://doi.org/10.1089/cyber.2009.0445

Hook, J. N., Hall, T. W., Davis, D. E., Van Tongeren, D. R., & Conner, M. (2021). The Enneagram: A systematic review of the literature and directions for future research. *Journal of Clinical Psychology, 77*(4), 865–883. https://pubmed.ncbi.nlm.nih.gov/33332604/

IPIP. (n.d.-a). *Permission*. International Personality Item Pool. https://ipip.ori.org/newPermission.htm

IPIP. (n.d.-b). *Translations*. International Personality Item Pool. https://ipip.ori.org/newTranslations.htm

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87*(4), 797–807. https://doi.org/10.1037/0021-9010.87.4.797

Kariyer Planım. (n.d.). *Kariyer profil testi*. http://www.kariyerplanim.com/

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19–40. https://psychology.psy.sunysb.edu/attachment/measures/content/maccallum_on_dichotomizing.pdf

Mavletova, A., & Couper, M. P. (2014). [Abandonment rates in mobile and desktop web surveys; GOR 2014 conference presentation]. https://cts.hse.ru/data/2014/04/03/1317876257/Mavletova_Couper_GOR_2014.pdf (reference to be verified)

McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality, 57*(1), 17–40. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x

O*NET Resource Center. (n.d.-b). *O*NET Tools Developer License*. https://www.onetcenter.org/license_toolsdev.html

Perkmen, S., & Tezci, E. (2015). [A 55-item Holland-based interest scale in pre-service teachers]. Balıkesir University open access. https://acikerisim.balikesir.edu.tr/items/d7bef6e3-02d7-4e88-91b1-38e45092a012/full (reference to be verified)

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research, 57*(3), 210–221. https://doi.org/10.1037/1065-9293.57.3.210

Poškus, M. S., Kairys, A., Liniauskaitė, A., & Žukauskienė, R. (2014). "Your NEO summary": Is it a valid way of providing feedback, or is its accuracy due to the Barnum effect? *Psichologija, 49*, 44–59. https://doi.org/10.15388/Psichol.2014.49.3694

Remetesting. (n.d.). *Test satış*. https://remetesting.com/Home/TestSatis

Revilla, M., & Höhne, J. K. (2020). How long do respondents think online surveys should be? New evidence from two online panels in Germany. *International Journal of Market Research, 62*(5), 538–545. https://www.sowi.uni-mannheim.de/media/Einrichtungen/gip/Bilder/Dokumente/revilla_hoehne__2020__-_online_survey_length.pdf

Somer, O., Korkmaz, M., & Tatar, A. (2002). Beş Faktör Kişilik Envanteri'nin geliştirilmesi-I: Ölçek ve alt ölçeklerin oluşturulması. *Türk Psikoloji Dergisi, 17*(50), 21–36. https://avesis.mcbu.edu.tr/yayin/9f580945-7e2e-48e8-9a28-d68cfc6c173e/development-of-five-factor-personality-inventory-bes-faktor-kisilik-envanterinin-gelistirilmesi-i-olcek-ve-alt-olceklerin-olusturulmasi

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143. https://pubmed.ncbi.nlm.nih.gov/27055049/

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of Big Five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718–737. https://doi.org/10.1037/0022-3514.94.4.718

Süerdem, A., & Erkök, B. (2016). Assessing the reliability and validity of a shorter version of RIASEC in Turkish. *SHS Web of Conferences, 26*, 01063. https://doi.org/10.1051/shsconf/20162601063

Traitify. (n.d.). *The science*. https://www.traitify.com/science

Truity. (n.d.). *TypeFinder personality test*. https://www.truity.com/test/type-finder-personality-test-new

Trustpilot. (n.d.). *16Personalities reviews*. https://ca.trustpilot.com/review/16personalities.com

University of Connecticut. (n.d.). [Adding prescriptive feedback to written personality feedback; doctoral dissertation, n = 132]. https://digitalcommons.lib.uconn.edu/dissertations/AAI3504792 (reference to be verified)

2. Theoretical Framework and Literature

This chapter gathers, in an ordered narrative, the measurement traditions on which the test rests and the literature that constrains its design. It first addresses the Five-Factor Model, the basic framework of personality measurement, and its public-domain operationalisation, the IPIP-NEO-120 (2.1–2.2); it then compares alternative scales together with their licence status (2.3) and reviews the Turkish adaptation literature (2.4). On the vocational side it examines Holland's RIASEC model and the O*NET Interest Profiler (2.5), work values (2.6) and chronotype measurement (2.7). The second half of the chapter is devoted to the pressures the product faces: type models and their critique (2.8), response quality (2.9), abandonment and duration in tests taken before sign-up (2.10) and the products on the market (2.11). Each section ends with an "Inference for this study" paragraph summarising what the literature says to the design of this work. The recommendations in these paragraphs are the authors' interpretation; sourced findings are given with citations in the rest of the text.

2.1 The Five-Factor Model and the domain/facet hierarchy

The Five-Factor Model (Big Five) summarises personality differences in five broad dimensions: Emotional stability (the reverse pole of Neuroticism), Extraversion, Openness to experience, Agreeableness and Conscientiousness. The model grew out of the lexical tradition, that is, from factor analyses of the personality adjectives in a language, and in this respect it has also been tested for Turkish: the Turkish lexical study conducted by Goldberg and Somer with 631 participants and 498 adjectives found variants of the Big Five and an additional Attractiveness factor (Goldberg & Somer, 2000). A reanalysis of the same data showed that six Turkish lexical factors corresponded closely to the HEXACO model (Wasti et al., 2008). At the cross-cultural level, in a study of 17,837 people in 56 countries and 28 languages the five-dimensional structure proved robust across regions; the T-scores of the Turkish sample (N = 206) lay between 47.9 and 52.7, that is, in a band close to the international mean (Schmitt et al., 2007).

The model is hierarchical: each broad domain consists of narrower facets. In the NEO PI-R tradition each domain has six facets, and the 30 facets in total give a finer-grained description of personality. The Conscientiousness domain, for example, divides into the facets self-efficacy, orderliness, dutifulness, achievement-striving, self-discipline and cautiousness. Which of the domain and facet levels is the "right" resolution is an extension of an old tension in measurement between bandwidth and fidelity: broad scales predict a large number of behaviours moderately, narrow scales predict specific behaviours with greater accuracy.

The literature supports the practical value of the facet level. Paunonen and Ashton (2001) found that narrow scales explained more variance than the broad factors in 40 behavioural criteria and added explanation over and above the factors; aggregation loses criterion-specific variance (as cited in Briley & Tucker-Drob, 2012). These findings are accompanied by a methodological caution: facet gains can be inflated by reliability differences and scale overlap, and it is therefore recommended that domain variance be separated from facet-specific incremental variance. Scale length is also a critical variable.

Credé et al. (2012) showed in 437 employees and 395 students that longer scales almost always yielded higher criterion correlations; the relationship between Conscientiousness and task performance was about .6 with an 8-item scale and fell to about .2 with a single-item scale.

The price of the facet level is reliability. As the number of items per facet decreases, internal consistency falls; 4-item facets are adequate for research but remain borderline for individual decisions (Johnson, 2014). The test review model of the European Federation of Psychologists' Associations (EFPA) likewise notes that when the reliability of subscales is about .70, the domain scales composed of them can reach .90, and requires that the report be no more specific than the level of detail the scale measures (EFPA, 2025).

Inference for this study. The facet level is not a luxury but the raw material of a person-specific report: a report written from domain scores is obliged to say the same sentences to everyone in the same domain band. However, because the reliability of facet scores is markedly lower than that of the domains, we adopted two principles: report headings are built at the domain level and the detail comes from facet patterns; and facets are treated not as percentiles but as three bands, low, medium and high (ADR-026). As of 22 November 2026 the report does not contain a separate breakdown of the thirty facets; the Personality section shows the five domains (`Front-end/src/components/test/report/pages/Personality.tsx`). The fact that the Turkish sample came out close to the international mean suggests that an international reference would not create a large deviation; personality percentiles today rest on such a reference, the published international sample of the IPIP-NEO-120 (Johnson, 2014; ADR-082). Since there is no evidence of measurement invariance, carrying this norm over to the Turkish adaptation is an assumption, and a Turkish norm remains an open task (ADR-029).

2.2 The IPIP and the IPIP-NEO-120

The International Personality Item Pool (IPIP) is a pool of personality items created under Goldberg's leadership and published in the public domain. The IPIP permission statement states explicitly that permission is automatically granted to use the items, scales and inventories for any purpose, commercial or not, and that there is no need to ask for permission separately (IPIP, n.d.-a). This provision made it possible to rebuild the facet structure of the copyrighted NEO PI-R with public-domain items. The IPIP-NEO-300 derived from the pool contains 10 items per facet, and its facet alphas range from .71 to .88 (IPIP, n.d.-c).

The IPIP-NEO-120 was developed by selecting 4 items for each facet from the 300-item form (Johnson, 2014). The development sample consisted of 21,588 people from the internet; the form was tested in the Eugene-Springfield community sample and in two large internet samples of 307,313 and 619,150 people. In the sample of 619,150, domain alphas range from .81 (Openness) to .90 (Neuroticism and Conscientiousness) and facet alphas from .63 to .88; all facets but three are at .69 or above. The weakest facets were O6 Liberalism (.63), O3 Emotionality (.65) and C3 Dutifulness (.67). The correlation of the facet scales with their NEO PI-R counterparts lies in the .53–.80 band, and for the domain scores it is .94 with correction for unreliability. Johnson's own assessment draws a binding limit for the product's language: the facet scales have sufficient reliability for research, but probably should not be used to make important decisions about individuals.

An independent reanalysis largely confirms this picture. In an online sample of 320,128 people in the United States the mean facet alpha was .78, and four facets fell below .70: O3 Emotionality, C3 Dutifulness, O6 Liberalism and A4 Cooperation (Kajonius & Johnson, 2019). The fit of the second-order confirmatory factor analysis is "at best tolerable"; the loading of O6 on the Openness domain is only .35, and the loadings of the A5 Modesty items on the domain remained at very low values of .05–.09. The authors explain this by social desirability.

Two structural properties of the form become important in adaptation. First, in the paper-and-pencil community sample the facet alphas are markedly lower than in the internet samples; in the sample of 501 people C3 was .47 and A4 was .56, and the mean facet alpha, which was .80 in the 300-item form, dropped to .68 in the 120-item form (Johnson, 2014). Second, in selecting items Johnson prioritised alpha rather than keying balance: 65 of the 120 items are positively keyed and 55 negatively keyed, and in some facets all items are keyed in the same direction. A script using reverse-keyed item pairs was provided separately to correct for acquiescence bias.

On the norm side, Johnson's open data repository is a usable source. Twomey and Johnson derived IPIP-NEO-300 norms for the 14–17, 18–29 and over-30 age groups from the United Kingdom and Ireland subset of this repository (N = 18,591) and described how the method can be applied to other countries; the representativeness of the sample is limited, and it consists of self-selected participants (Twomey & Johnson, 2022). The number of rows from Turkey in this repository is not known in the notes. Johnson's IPIP-NEO-120 open data (619,150 people, https://osf.io/wxvth/) has, since 22 November 2026, also been the source of the personality percentiles in Rehnüma (ADR-082).

Inference for this study. When licence and facet coverage are considered together, the IPIP-NEO-120 is the only clean option for a commercial Turkish product; the public-domain provision applies to the items, and Johnson's norm tables and software must be evaluated separately. For the percentiles we used not the published norm tables but the open data of the same study: separate tables were computed from the responses of 619,150 people for the 60 core items of stage 1 and the 120 items of stage 2; the mid-rank percentile of every possible total score is stored, and no normal distribution is assumed (ADR-082; `Pusula-API/app/Profile/PublishedNorms.php`). Of the 120 items, 117 matched the original key; because the three rewritten O6 items are not present in the data, no percentile is given for the O6 facet and the Openness domain is normed with the remaining original items. The norm comes from volunteers who completed the English form online, is not broken down by age or sex, and carrying it over to the Turkish adaptation is an assumption. Two weaknesses of the form are inherited by the adaptation and must be addressed explicitly in the design: weak facets such as O3, O6, C3 and A4, and facets keyed in a single direction. Because 4 items per facet do not fit into a single session before sign-up, a core of 2 items per facet with completion after sign-up was envisaged; this structure is detailed in Chapter 3. The weak facets are expected to be weak in Turkish as well; reading the pilot results side by side with the original alphas is the only way to distinguish a translation error from the weakness of the source.

2.3 Alternative scales and their licences

The IPIP-NEO-120 is not the only option. Other instruments widely used in personality measurement offer different trade-offs in facet coverage, length and reliability. Table 2.1 compares these instruments.

Table 2.1. Comparison of Big Five and related personality scales.

Scale	Items	Structure	Reliability (source sample)	Status in Turkish
IPIP-NEO-300	300	5 domains × 6 facets, 10 items per facet	Facet α .71–.88	No published adaptation found
IPIP-NEO-120	120	5 domains × 6 facets, 4 items per facet	Domain α .81–.90; facet α .63–.88 (N = 619,150)	No published adaptation found
BFI-2	60	5 domains × 3 facets, 4 items per facet	Domain α ≈ .81–.89; facet α mean ≈ .74, lowest .56 (Czech and Danish samples)	Cemalcılar et al. (2021); internal-consistency differences in the community sample
HEXACO-100 / 60	100 / 60	6 domains × 4 facets (100); domains only (60)	TR long form domain α .72–.78, facet α .29–.77	Tatar (2018)
IPIP-50 (Big-Five Factor Markers)	50	5 domains, no facets	α .79–.87	B5KT-50-Tr, factor α .65–.79 (Tatar, 2017)
Mini-IPIP	20	5 domains, 4 items per domain	α ≈ .60 and above	No adaptation found in the notes
TIPI	10	5 domains, 2 items per domain	TR α .81–.86, retest .87–.89	Atak (2013)
Quick Big Five	30 adjectives	5 domains, 6 items per domain	TR ages 14–17 α .71–.81, retest .80–.87	Morsünbül (2014)

Note. Values are compiled from Johnson (2014), IPIP (n.d.-c), Hřebíčková et al. (2020), the Danish BFI-2 study (Vedel et al., 2021), Tatar (2017, 2018), Donnellan et al. (2006), Atak (2013) and Morsünbül (2014). The Turkish alphas of the TIPI are unusually high for two items per domain and should be read with caution.

The BFI-2 is the closest alternative to the IPIP-NEO-120. With 60 items and 15 facets it fits comfortably within the time budget, and with its equal numbers of true-keyed and reverse-keyed items it allows acquiescence bias to be controlled by within-person centring (Soto & John, 2017). In the Czech adaptation, with 1,733 people aged 15–26, six-month retest reliability averaged .86 for the domains and .80 for the facets; the 30-item and 15-item short forms capture about 91% and 77%, respectively, of the domain information of the full form. The HEXACO adds a sixth dimension called Honesty-Humility, and its measurement invariance has been tested in 16 languages; in samples including Turkish, configural and metric invariance were supported, while scalar invariance was generally not supported. The short scales (IPIP-50, Mini-IPIP, TIPI) yield domain scores but no facets; the internal consistency of the Mini-IPIP was found to be .60 and above across five studies (Donnellan et al., 2006).

For a commercial product the real differentiator is the licence. The permission statement on the BFI-2 form says that the test content may be reproduced without permission only for non-commercial research and educational purposes, and that every other use requires written permission from the authors and the publisher. The HEXACO website states that the forms are free of charge only for non-profit academic research and does not allow even researchers to collect data through online survey sites or applications open to the general public (HEXACO, n.d.). Similar restrictions exist for constructs outside personality: the VIA terms of service limit the content to personal and non-commer-

cial use (VIA Institute on Character, n.d.); the Grit-S (Duckworth, n.d.) and the Satisfaction with Life Scale (Diener, n.d.) are open only for non-commercial purposes. Table 2.2 summarises the licence terms of the scales that entered the product decision.

Table 2.2. Licence status of the scales considered, with respect to commercial use.

Scale or source	Verdict	Condition and note
IPIP items (IPIP-NEO-120, IPIP-50, Mini-IPIP, IPIP-VIA)	Public domain	Use without permission for any purpose, including commercial; norm tables and software are separate
O*NET Interest Profiler, WIL, WIP	Free with attribution, conditional	Developer License: worldwide, royalty-free; if modified or used in part, a validation study before release and a mandatory notice (§3(b)); endorsement may not be implied
O*NET Database (occupations, Work Styles, value profiles)	Free with attribution	CC BY 4.0; attribution and statement of changes
BFI-2	Non-commercial only	Other use requires written permission from John and Soto
HEXACO-PI-R	Academic only	Online administration open to the general public is prohibited
VIA-IS and derivatives	Non-commercial only	Commercial use requires a separate agreement; modification subject to written permission
Grit-S	Non-commercial only	Duckworth laboratory scales
SWLS	Non-commercial only	Diener and co-authors
BSCS, CAAS, NFC	Unclear	No explicit licence statement found; written permission required
CSE	Unclear, favourable	"Nonproprietary" only in a secondary source
MEQ, rMEQ	Unclear	No explicit commercial licence text found
MCTQ	Permission required	Developer's permission; the MIT licence of the R package does not cover the questionnaire
MBTI items and trademarks	Copyrighted and trademarked	Items may not be reproduced without written permission; "MBTI" and "Myers-Briggs" are registered trademarks
RHETI (Enneagram)	Copyrighted	Publisher's permission

Note. Verdicts are compiled from the relevant licence pages and from research notes (IPIP, n.d.-a; O*NET Resource Center, n.d.-a, n.d.-b; BFI-2 form; HEXACO, n.d.; VIA Institute on Character, n.d.; Duckworth, n.d.; Diener, n.d.; The Myers-Briggs Company, n.d.). The table is not legal advice; scales with an "unclear" verdict should not be used in a commercial product without written permission.

Inference for this study. Psychometrically, the BFI-2 is an attractive alternative on account of its number of facets, its keying balance and its short forms; however, commercial use is subject to a written licence and its Turkish form is the product of a separate team. The HEXACO is in effect closed to web applications open to the general public. Because the short scales do not yield facets, they do not meet the need for a person-specific report. For these reasons we chose the IPIP-NEO-120; the BFI-2 can nevertheless be used as a design template, not for its items but for its structure, in particular its equal keying balance. The licence table also explains why sections offered by competing reports, such as character strengths, grit and life satisfaction, cannot be imitated directly: the same narrative has to be built either with public-domain IPIP counterparts or with tendencies derived from the facets (ADR-025).

2.4 The Turkish adaptation literature

Turkish personality measurement has two main branches: instruments derived from the IPIP pool and direct adaptations of international scales. At the centre of the first branch stands the work of Somer, Korkmaz and Tatar. The authors translated 924 IPIP items into Turkish and from this pool developed the Five Factor Personality Inventory (Beş Faktör Kişilik Envanteri, 5FKE), containing 220 items, 5 factors and 17 subdimensions (Somer et al., 2002). The IPIP translations page lists this translation but notes that its accuracy has not been checked by anyone associated with the IPIP project (IPIP, n.d.-b). The short form of the 5FKE consists of 76 items, 6 social desirability items and 3 control items and has a norm group of 5,239 people; however, the inventory does not follow the 30-facet structure of the IPIP-NEO and may be a commercial scale. In the Turkish translation of the 50-item IPIP form (B5KT-50-Tr), factor internal consistencies ranged from .65 to .79 and the five factors explained about 35.6% of the total variance (Tatar, 2017). The search found no thesis or article on a Turkish adaptation of the IPIP-NEO-120 or the IPIP-NEO-300; this is not evidence of absence but a record of not having found one.

Adaptations of international scales are more scattered. The Turkish BFI-2 was tested in university students and in a nationally representative community sample aged 18–35; in the students the psychometric properties matched the original, whereas in the community sample the factor structure was replicated but notable differences in the internal consistency of the domains and deviations in some trait–outcome relationships were observed (Cemalcılar et al., 2021). In the Turkish long form of the HEXACO the alphas of the 24 subdimensions varied between .29 and .77, and 15-day retest reliability was .67–.80 (Tatar, 2018). As a Turkish example working with adolescents, the Quick Big Five yielded alphas of .71–.81 and retest reliabilities of .80–.87 at a two-week interval in 478 high-school students aged 14–17 (Morsünbül, 2014). The Turkish adaptation of the TIPI, in turn, can be used as a short, free convergent validity criterion (Atak, 2013).

On the vocational interest side the Turkish literature is weaker still. Kuzgun's Self-Assessment Inventory (Kendini Değerlendirme Envanteri) is widespread in school guidance, but with its 11 interest areas it is not a one-to-one RIASEC measure. Atlı and Keldal's Vocational Personality Types Inventory was developed with 1,129 high-school students (Atlı & Keldal, 2017). The Turkish short form of the Personal Globe Inventory (PGI-S-T) showed strong fit to both the eight-type and the RIASEC circular model in high-school and university students (Vardarlı et al., 2017). By contrast, it was concluded that the 41-item short scale developed for business professionals required substantial revision, after experts criticised the word-for-word translation, the machinery items were perceived as men's work and the first confirmatory analysis showed poor fit (Süerdem & Erkök, 2016). In a 55-item scale administered to pre-service teachers, the correlation and multidimensional scaling results contradicted the Holland hexagon (Perkmen & Tezci, 2015). In these two studies the type assignment of the Enterprising and Conventional items shifted in translation.

For the adaptation method, the international framework is the second-edition guidelines of the International Test Commission (ITC). The guidelines recommend intellectual property permission, at least two independent forward translations with reconciliation, cognitive pretesting, pilot item analysis with a modest sample (for example 100 people), 300 or more people for factor structure and at least 200 people per version for differential item functioning; they stress that back-translation alone is not sufficient, because it can produce "a target text that is easy to back-translate but poor" (ITC, 2017). In survey research the counterpart of the same workflow is the TRAPD method (translation,

review, adjudication, pretesting, documentation), and the European Values Study uses this method as the best practice currently available (European Values Study, n.d.). There is an additional risk for adolescents: in data from ages 10–20 the main threat is not word comprehension but acquiescence bias, which varies widely among young people and strongly affects the psychometric properties of scales (Soto et al., 2008).

Inference for this study. Because no published Turkish adaptation of the IPIP-NEO-120 or the O*NET Interest Profiler exists, the adaptation work has to be done from scratch; the item wordings of the 5FKE and the B5KT-50-Tr can be used as a reference during translation but should not be taken over verbatim. The Turkish RIASEC experience shows that the Enterprising and Conventional items, and Realistic items carrying gender stereotypes, require particular attention. The ITC guidelines recognise no exception for a product whose translation is done by its own developer; the minimum acceptable route is a second independent translation, a reconciliation record and cognitive interviews with adolescents. The existence of the Turkish BFI-2 and the TIPI-TR, for their part, provides ready benchmark scales for the convergent validity analysis at Gate 3 (ADR-029). The fact that the Turkish Quick Big Five reaches the .70 band in adolescents suggests that an adolescent pilot with short statements free of negation can aim for this band.

2.5 Holland's RIASEC and the O*NET Interest Profiler

Holland's theory of vocational personalities groups interests into six types: Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E) and Conventional (C). The types are arranged on a hexagon in order of adjacency; adjacent types resemble one another and opposite types are distant from one another. A person's profile is summarised by a code made up of the letters of the three highest types, and this code is matched with occupational environments coded in the same way. The theory also defines two further properties of the profile: elevation (the sum of the six scores, that is, the general level of interest) and differentiation (the spread between the highest and the lowest score). High profile elevation has been found to be associated with high Openness to experience and Extraversion (Fuller et al., 1999). Research support for the concepts of differentiation and consistency, by contrast, is contradictory; cosine-based congruence indices computed from the whole profile have been found superior to the traditional definitions, and the standard deviation of the person's six scores (amplitude) has been proposed for differentiation.

The predictive power of interests stands apart from personality. In a meta-analysis covering 60 studies and about 568 correlations, interests were found to be related to performance and persistence in work and academic settings; indices of congruence between person and environment yielded stronger correlations than interest scores alone (Nye et al., 2012). Interests measured at the end of high school predicted seven of nine life outcomes 10 years later over and above intelligence and the Big Five; they were stronger than the Big Five for work and relationship outcomes and weaker for health outcomes (Stoll et al., 2017). Interests predict career and educational choice and, given their stability across the lifespan, are supported as separate dispositions rather than as extensions of basic personality traits (Rounds & Su, 2014). According to a longitudinal meta-analysis, interest scores decline in early adolescence and increase in late adolescence, and stereotypical gender differences peak in early adolescence (Hoff et al., 2018); in a review of 9,494 people aged 13–60 in Iceland, the RIASEC structure fitted Holland's model except in the youngest age group (Einarsdóttir & Rounds, 2020).

The O*NET Interest Profiler is a RIASEC inventory that is matched to the occupational database of the U.S. Department of Labor and offered under a developer licence. It comes in three lengths. The 180-item long form, with 30 items per scale, yields α .95–.97. The 60-item Short Form (Short-IP), with 10 items per scale, yields α .78–.87 (mean .81) and retest reliabilities of .78–.86 in a stability sample of 132 people; in the randomisation test of the RIASEC circular order the Correspondence Index is .69 (Rounds et al., 2010). Of the Short-IP development sample, 9.6% were aged 18 or under; 26.7% of the student subgroup were high-school students. The response scale has five points from "strongly dislike" to "strongly like", and its midpoint carries the meaning "unsure". The 30-item Mini-IP is a subset of the Short-IP: items were selected according to the item response theory discrimination parameter, and 5 items per scale was deliberately set as a length that would not lower reliability too far. In the development data α was .70–.75, and in the validation sample of 575 people .74–.81; the report states explicitly that the Mini-IP is suitable for embedding within a longer multi-measure questionnaire (Rounds et al., 2016). Table 2.3 compares open or widely used RIASEC instruments.

Table 2.3. Comparison of RIASEC measurement instruments.

Instrument	Items	Per scale	Reliability	Licence
O*NET IP long form	180	30	α .95–.97	O*NET Tools licence
O*NET Short-IP	60	10	α .78–.87; retest .78–.86	O*NET Tools licence
O*NET Mini-IP	30	5	α .70–.81; primary code with Short-IP κ = .73	O*NET Tools licence
RIASEC Markers	2 × 48	8	Numerical value could not be verified	Said to be public domain; not confirmed
Strong Interest Inventory	244 (current)	—	General theme scales α .90–.95	Commercial
Self-Directed Search	—	—	"Rule of 8" rests on the standard error of measurement	Commercial

Note. Values are compiled from Rounds et al. (2010, 2016), the SDS handbook (PAR, n.d.) and secondary sources in the research notes. The Strong and the SDS are listed for comparison.

In profile interpretation the most fragile output is the primary code. Although four established RIASEC inventories correlated highly at the scale level, they frequently gave the same person different primary codes (Chu et al., 2022). Even between two lengths of the same item pool, between the Mini-IP and the Short-IP, primary-code agreement remains at κ = .73; although the median within-person profile correlation is .95, the first letter changes in roughly one person in every four or five (Rounds et al., 2016). The Self-Directed Search (SDS) manages this problem with the "Rule of 8": on the basis of the standard error of measurement it treats differences smaller than 8 points as negligible. Glavin and Savickas (2011), in 2,397 undecided undergraduates, found the hit rate for choice of major to be 55% (κ = .38) when the primary-code difference was 8 or above and 40% (κ = .21) when it was below.

The scoring format is a second design question. Normative measures (Likert scales in which each item is rated independently) rest on between-person comparison, ipsative measures (rankings with a fixed total, or forced choice) on within-person differences. Ipsative scores are less affected by social desirability among options of equal desirability, but fully ipsative scores violate the assumptions of classical test theory, which distorts reliability and construct validity estimates (Morillo et al., 2019). Ipsativity is a property of the scoring method, not of the format: forced-choice items can be scored normatively with Thurstonian item response models. In a classic comparison in which values were

measured by ranking and by rating, the two methods gave similar results for the aggregate order of preferences and different results for the latent structure (Alwin & Krosnick, 1985). In personality inventories, the operational validity of quasi-ipsative formats has been found to be markedly higher than that of fully ipsative formats.

The third concept is expressed interest: the preference a person states directly about himself or herself ("what I would most like to do is this"). The Daydreams section of the SDS scores this information separately and interprets it side by side with the measured summary code as a distinct aspiration (Daydreams) code. In Holland's own example in the handbook the measured code came out as AEI and the aspiration code as SAE; this difference was read not as an error but as different facets of personality (PAR, n.d.). Expressed interests have been reported to predict current choice in university students better than measured interest scores (Purdue University, n.d.). However, no published precedent could be found that blends expressed interest with the measured score in a specified proportion.

The licence conditions determine the design with the same weight as the theoretical discussion. The O*NET career exploration tools are offered under two options: CC BY-ND 4.0 for copying without modification, and the Tools Developer License for modification and translation (O*NET Resource Center, n.d.-a). The developer licence is worldwide, royalty-free and irrevocable, and does not prohibit commercial use. On the other hand, if only part of the material is used or it is modified in any way, then before the product is released a validation study consistent with the standards for educational and psychological testing must be carried out, showing that the results are valid for the intended purpose, audience and end user; in addition, the opening screen must display the English text stating that USDOL/ETA has not approved, endorsed or tested the modifications, together with a link containing the phrase "modifications and disclaimers" (O*NET Resource Center, n.d.-b, §3(b)). The licence does not define a minimum sample size. The O*NET database that feeds the occupation matching, for its part, is licensed under CC BY 4.0 and requires only attribution (O*NET Resource Center, n.d.-d).

Inference for this study. For interest measurement the single O*NET pool can be used at two lengths: the 30 items of the Mini-IP before sign-up and the remaining 30 items of the Short-IP after sign-up; no item is asked twice. Because going beyond 60 items raises α only to .95–.97, the additional time should be spent on other axes. The presence of high-school students in the Short-IP development sample suggests that a separate item pool is not mandatory for an audience aged over 15, and that age-neutral items and separate interpretive text may be sufficient. Because of the fragility of the primary code, the report should show not a single three-letter code but the whole profile, the elevation and the differentiation; if the difference between the first two areas lies within measurement error, it should say so as "neck-and-neck". Expressed interest should not be mixed into the score; like the aspiration (Daydreams) code in the SDS it should stand beside the measured result as "what you said", and should only break ties that lie within measurement error (ADR-026). In the current implementation expressed interest is still asked and only breaks these ties (`Pusula-API/app/Scoring/V4/ExpressedInterest.php`); the "what you said and what was measured" comparison in the Interests section of the report, however, was removed on 22 November 2026. Because translation and shortening trigger the licence's validation obligation, the test cannot be presented as though its validity had been demonstrated before this study is completed, and a phrase such as "O*NET approved" cannot be used.

2.6 Work values and the Work Importance Locator

Work values describe what a person finds satisfying at work and add a third axis to the questions "what do you like doing" (interests) and "what are you like" (personality). The theoretical basis is Dawis and Lofquist's Theory of Work Adjustment: job satisfaction is a function of the correspondence between the person's needs and the rewards the work environment provides, and the latent dimensions underlying needs are called values. According to the theory, needs with high scores are important for satisfaction and those with low scores have no effect; for example, the level of independence in a job markedly affects the satisfaction only of people with a high need for independence (McCloy et al., 1999a).

The O*NET work values instruments are derived from the Minnesota Importance Questionnaire and measure 21 needs grouped under six values: achievement, independence, recognition, relationships, support and working conditions. The computerised Work Importance Profiler (WIP) has the 21 statements ranked in 21 blocks of five; this arrangement reduces 210 paired comparisons to 21 blocks. The "absolute zero" section that follows asks about the absolute importance of each need in order to compensate for the ipsativity of the ranking. The paper-and-pencil Work Importance Locator (WIL), in turn, has 20 cards sorted into five piles of four cards each; in the pre-pilot the mean time was 14 minutes, and pile and calculation errors were frequent (McCloy et al., 1999a). Both instruments have been retired, but the Tools licence continues to cover them (O*NET Resource Center, n.d.-c). In their place O*NET today points to CareerOneStop's Work Values Matcher, a tool that has 20 statements sorted into columns; the column structure could not be verified from a primary source.

The reliability of these instruments is weak, and the source reports acknowledge this openly. Because of ipsative scoring the median α in the WIL is .20; retest reliability is .26–.62 (median .42) at the need level and, corrected, .33–.61 (median .52) at the value level. The top value was the same on the two occasions in 62% of cases, and the top value at the first measurement was among the top two at the second in 80% of cases (McCloy et al., 1999a). In the WIP development report, 4–8-week retest reliability at the value level is .59–.66 and alpha ranges from .50 to .86; the first-ranked value remains first at the second administration in 62–70% of cases and within the top two in 88% (McCloy et al., 1999b). The fact that O*NET carries only the two highest values into the score report is a design response to this weakness. In scoring, different multipliers are applied to the values and each value is scaled to the range 6–30.

Evidence on whether values contribute anything beyond interests and personality is limited but positive. In the O*NET-based study of "whole-person" fit, each of five fit domains, namely interests, values, knowledge, skills and personality, made a unique contribution to subjective career outcomes, and interests were generally the strongest predictor (Liu et al., 2025); the incremental validity coefficients in the published version could not be verified in the notes, and the findings were taken from the preceding thesis. In a meta-analysis of person–environment fit (172 studies, 836 effect sizes), person–job and person–organisation fit were found to be related to attitudes, performance, turnover and strain (Kristof-Brown et al., 2005). This evidence is confined to job satisfaction; no study could be found showing that work values predict everyday goal motivation or leisure, health and financial behaviour (evidence grade: none).

Inference for this study. Because they directly answer the question "why am I unhappy in this job", especially for working adults, work values are an axis with a high report value per minute; because O*NET's occupational value profiles are CC BY 4.0, they can also be tied into occupation matching. A short ranking of the six values gives a rough "top two values" but loses the needs beneath each value; a better middle way is a card sort of 20 statements in five columns (ADR-026). The source's own stability data require that only the top two values be emphasised in the report and that the full ranking of the six values not be presented as a definitive result. The absence of evidence that values predict everyday behaviour likewise shows that this variable should remain in the report narrative and should not be a condition of any widget rule on the dashboard (ADR-033). In the product copy, saying "adapted from the O*NET work values framework" rather than using the name "O*NET Work Importance Locator" like a brand is more in keeping with the licence's prohibition on implying endorsement.

2.7 Chronotype measurement

Chronotype is a circadian characteristic describing the hours of the day at which a person is predisposed to be alert, productive and sleepy. There are two traditions in its measurement. The first is the preference tradition: Horne and Östberg's 19-item Morningness-Eveningness Questionnaire (MEQ) and its 5-item abbreviation, the rMEQ (Horne & Östberg, 1976; Adan & Almirall, 1991). These instruments ask at what hours the person prefers to get up, to work or to sit an examination, and produce a score together with morning, intermediate and evening classes. The second is the behaviour tradition: the Munich ChronoType Questionnaire (MCTQ) asks about actual sleep times on work days and free days and computes a continuous measure of chronotype by correcting the midpoint of sleep on free days for sleep debt (MSFsc). The MCTQ has collected about 300,000 online entries over more than 15 years and has been translated into 13 languages (Roenneberg et al., 2019).

Chronotype changes markedly with age. In a representative U.S. sample (N = 53,689) the weekend midpoint of sleep was 04:52 at age 20 and moved back to 03:02 at age 60; men were later than women until about age 41 and slightly earlier thereafter (Fischer et al., 2017). Chronotype is moderately stable over a seven-year follow-up (r = .56), and the mean shift is small (Druiven et al., 2020). In Turkish student samples classified by the MEQ, the intermediate type is the overwhelming majority at 68–83%.

The relationship of chronotype to personality is known at the meta-analytic level. In a meta-analysis covering 44 independent samples and 16,647 people, morningness was found to be related most strongly to Conscientiousness (ρ = .37); this relationship persisted when the other domains were controlled (β = .32), a small negative relationship was observed with Neuroticism, Agreeableness was unrelated, and personality explained 10.9–16.4% of the variance in chronotype (Lipnevich et al., 2017). An earlier meta-analysis gave the Conscientiousness–morningness relationship as r = .29 (Tsaousis, 2010). These magnitudes show that chronotype is not a by-product of personality but a variable that overlaps with it yet is distinct.

Turkish adaptations exist for two instruments. In the Turkish reliability study of the MEQ, 618 students and staff were measured twice at an interval of 15–20 days, and no significant difference was found between the means (Pündük et al., 2005). The Turkish rMEQ showed good fit to a unidimensional structure in 875 university students (CFI = .990, RMSEA = .047); α = .706 and ω = .740, and

sensitivity and specificity range from 83.3% to 92.7% for the morning type and from 86.3% to 87.3% for the evening type (Uygur et al., 2024). The correlation between the Turkish rMEQ and the MCTQ adapted for shift workers is around $r = -.47$; that is, the preference and behaviour measures overlap moderately. The agreement of a single-item self-assessment with MSFsc, by contrast, remained at a weighted $\kappa = .39$.

The licence status is the most unclear in this area. The MEQ and the rMEQ have been published in peer-reviewed journals and are widely reproduced, but no explicit commercial licence text could be found in the notes. For the MCTQ, the ATS list of sleep questionnaires indicates developer permission; the MIT licence of the R package written for MCTQ computation covers only the software code, not the questionnaire itself (American Thoracic Society, n.d.; rOpenSci, n.d.). Moreover, because the MCTQ is computed from free-form clock-time responses and is not a short instrument, it is not suitable for a web test taken before sign-up.

Inference for this study. Chronotype is an axis whose contribution to a career report is limited but which carries a synergy specific to Rehnüma: the sleep plan, training time and habit timing can be tied to it directly. The Turkish rMEQ, with five items and about one minute, offers good value per minute, but its licence is unclear. For this reason, until written permission arrives, we used three short preference questions in place of the scale, and these are not presented as a psychometric scale (ADR-026). The moderate relationship between chronotype and Conscientiousness shows that the two variables should not be used in place of one another but read together and as separate rules: the hour should be derived from rhythm, and frequency and intensity from self-discipline. The shift of chronotype with age and its moderate stability also suggest that a single measurement should not be considered valid for years, and that once the user's actual sleep records accumulate, these records should take precedence over self-report. The sleep times accumulating in Rehnüma's habit module may be a sounder basis for the midpoint-of-sleep calculation of the MCTQ tradition than subjective items.

2.8 Type models: the MBTI and Keirsey

Type models assign people to a limited number of qualitative categories. The most widespread is the Myers-Briggs Type Indicator (MBTI), derived from Jung's theory of psychological types. The MBTI defines four dichotomous preferences: Extraversion–Introversion (E–I), Sensing–Intuition (S–N), Thinking–Feeling (T–F) and Judging–Perceiving (J–P); their combination produces 16 types and four-letter codes (for example INFJ). Keirsey's temperament classification is a second layer built on top of these codes. Keirsey defined the four temperaments theoretically in 1978 and referred to them by names taken from Plato: Guardian (SJ), Artisan (SP), Idealist (NF) and Rational (NT). The first division is between concrete and abstract (S/N); the second division falls on J/P on the concrete side and on T/F on the abstract side.

Relationship to the Big Five. The most robust finding about type models is that the four MBTI indices measure four Big Five dimensions. McCrae and Costa (1989) showed, in 468 people aged 19–93 with NEO-PI self-reports and peer reports, that the MBTI measures four relatively independent dimensions and that these dimensions correspond to Extraversion, Openness to experience, Agreeableness and Conscientiousness; Neuroticism has no counterpart in the MBTI. The pattern was replicated with the NEO PI-R in 900 people (Furnham et al., 2003). Table 2.4 gives the correlations.

Table 2.4. Correlations between the MBTI scales and the Big Five dimensions (McCrae & Costa, 1989).

MBTI scale	Extraversion	Openness	Agreeableness	Conscientiousness	Neuroticism	Match
E–I	**−.74**	.03	−.03	.08	.16	Extraversion
S–N	.10	**.72**	.04	−.15	−.06	Openness to experience
T–F	.19	.02	**.44**	−.15	.06	Agreeableness
J–P	.15	.30	−.06	**−.49**	.11	Conscientiousness

Note. N = 468 (267 men, 201 women). Signs depend on the direction of the MBTI continuous score; the I, N, F and P poles receive high scores. The values are taken from a secondary account and could not be compared one-to-one with the original table (research note `Kişilik testi doğrulama protokolü/turetilmis_katmanlar.md`, §2). Bold values are the main match of each scale.

Two conclusions follow from the table. First, the E–I and S–N scales are almost a renaming of a single Big Five dimension ($|r| \approx .72$–$.74$). Second, the link of the T–F and J–P scales to their counterparts is weaker: the relationship of T–F with Agreeableness remains at .44 and also takes a share from Extraversion; J–P is related, besides Conscientiousness, to Openness at the .30 level. These two letters are structurally noisier than the others. As a counter-view, one systematic review found the reliability of the E–I, S–N and J–P subscales to be .75 and above and that of T–F to be .61; a 25-year synthesis published on myersbriggs.org, for its part, reported high internal consistency while recording that construct validity and retest designs were absent from the literature of that period.

Dichotomisation. The theoretical claim of the MBTI is that on each axis people belong to one of two discrete groups. This claim is not supported by the data: McCrae and Costa (1989) found no support for the view that truly dichotomous preferences or qualitatively different types are measured, and the score distributions are not bimodal but peaked at the centre. The bimodal curves reported earlier were shown to be an artefact of the scoring software. Cutting a continuous distribution at its midpoint means that the people close to the cut, who form the majority, change letter with a small measurement error. In the methodological literature, splitting a quantitative variable in two is described as a practice that is "rarely defensible and often misleading", and it loses information about individual differences (MacCallum et al., 2002). There is also a concrete threshold relating to measurement error: assuming a reliability of .80, a score with $|z| \leq .78$ cannot be said at the 5% level to differ from the mean (Gerlach et al., 2019).

Retest instability. The direct consequence of dichotomisation is instability at the type level. Pittenger (1993) reported that 39–76% of those retested after five weeks received a different type (this range is a secondary account; it could not be verified in the primary source); a later assessment of his reports that at a five-week interval about half were classified differently on at least one scale (Pittenger, 2005). Critics note that these data rest largely on studies dating from 1962–1979. According to another account, over nine months about 83% of the individual dichotomous categories remained stable while only half of the people kept the same four-letter type. It is also reported that the 1991 assessment of the U.S. National Academy of Sciences found only the I–E scale strong and did not see sufficient research to justify use in career counselling. Stein and Swan (2019) concluded that the "true type" theory of the MBTI does not agree with known data, lacks testability and contains internal contradictions.

Keirsey. The evidence base of the temperament layer is thinner even than that of the MBTI. A strong positive correlation was found between the Keirsey Temperament Sorter-II and the MBTI in 203 university students, but it was stated that there was little information supporting the validity of the KTS-

II (Kelly & Jugovic, 2001). In a master's thesis the four temperaments did not discriminate dispositional optimism and explained only 6–8% of the variance in grade point average. No peer-reviewed study reporting direct correlations between the Keirsey temperaments and the Big Five could be found. By inference from Table 2.4, the temperaments correspond approximately to the following Big Five combinations: SJ ≈ low Openness and high Conscientiousness, SP ≈ low Openness and low Conscientiousness, NF ≈ high Openness and high Agreeableness, NT ≈ high Openness and low Agreeableness. Because the two second divisions rest on different factors, the partition is asymmetric, and there is no evidence that the four groups form empirical clusters in Big Five space (inference; evidence grade: none).

An evidence-based alternative: data-driven types. The question of whether types exist in Big Five data has been taken up again with large samples. Gerlach et al. (2018) showed, with more than 1.5 million people in four web data sets, that most of the common clustering solutions were spurious and that at least four robust types remained: average, role model (low Neuroticism, high on the other dimensions), self-centred (low Openness, Agreeableness and Conscientiousness) and reserved (low Neuroticism and low Openness). While four types separated in the 120-item and 100-item forms, only two could be separated in the 44-item form. In their subsequent reply the authors wrote that assigning the individual to a type was not the aim, that they agreed with the finding that only 42% of participants fell into a single type, and that they supported the call to avoid typology in practice (Gerlach et al., 2019). The way the clusters were tested against the null model has also been criticised; it has been shown that a skewed, unimodal distribution without cluster structure can produce "significant" clusters under the same criterion (Katahira et al., 2020).

The standard for category reporting. The International Test Commission's guidelines on computer-based testing require that, if reports assign test takers to categories such as "introverted type", the accuracy of the classification system be documented and it be explained from which score the report statements are derived (ITC, 2005, Guidelines 25.3 and 25.4).

Inference for this study. It is defensible to say that the four-letter code "measures something" at the axis level, because each axis is a renaming of a Big Five domain; but this shows precisely that the code should be derived from the Big Five and that measuring it with separate items is unnecessary. For this reason the code was defined as a secondary "trait code", deriving E/I from E, N/S from O, F/T from A and J/P from C; the clarity of each letter and its closeness to the boundary ("balanced") are shown, and a note that it is not related to the MBTI is mandatory (ADR-026). The structural fragility of the T–F and J–P letters should be stated openly in the report. Temperament, as the product of two axes, is only a narrative frame; because no predictive power specific to temperament has been shown, it cannot be a condition of any product decision (ADR-033). For the showcase label, the Gerlach clusters are the only evidence-based starting point, but this basis is limited: the clusters describe not the person but regions of density, do not separate in a short form and have negative names. For this reason the archetypes should be renamed with positive names, the person should be shown the several nearest archetypes with similarity percentages rather than a single type, and the assignment made at the stage before sign-up should be marked as "preliminary". The ITC requirement on classification accuracy turns measuring and publishing the retest stability of these layers into a product obligation (Chapter 4).

2.9 Response quality

In unproctored web tests some participants respond without reading the items or without attending to their content. In the literature this behaviour is called careless or insufficient-effort responding, and it takes two basic forms: straightlining (consecutive repetition of the same response, longstring) and random responding. To these are added response styles, in particular acquiescence bias (the tendency to say "agree" irrespective of content).

Prevalence. In student samples about 10–12% of participants responded carelessly (Meade & Craig, 2012); in summaries of the literature the rates vary from 3% to 46% depending on sample and method. For this product the most relevant evidence comes from adolescents: among 12,578 students (aged 15–21) in the final year of Flemish secondary schools, 14% of participants were flagged by at least one indicator under lenient thresholds and 37% under strict thresholds. Most of those flagged under the lenient threshold were caught by only a single method, and none of the five methods was common to all of them; the authors recommended that the indicators be used not as a final verdict but as a starting point for identifying suspect cases, and that caution be exercised in particular when giving individual feedback to students (Chauliac et al., 2023). Carelessness mostly arises not at the beginning of the test but in later sections, with fatigue and boredom (Welz & Alfons, 2023).

Indicators. There is no single correct threshold, and it is recommended that several indicators be used together in order to capture different forms of carelessness (Ward & Meade, 2023). The common indicators are of four kinds. Response time: later studies, citing the study by Huang et al. (2012), flag a page answered faster than two seconds per item. Longstring: no general rule can be given; one application takes as its threshold a run of identical responses equal to or longer than half the scale length, another proposal takes more than seven; a mixed order with reverse-coded items makes longstring more discriminating. Consistency indicators: psychometric synonym and antonym item pairs and odd–even consistency; however, in the 300 IPIP-NEO items the predetermined antonym threshold of $-.60$ yielded no pairs at all, and the threshold had to be lowered to $-.144$ (Brühlmann et al., 2024). Instructed-response items: attention-check items of the type "Select this option in this row". Mahalanobis distance captures only fully random responding and requires a large sample for a reliable covariance matrix.

Design of attention-check items. A study of 816 people showed that instructed-response items do not harm scale validity (Kung et al., 2018). However, the detail of the design changes the outcome: an instructed-response item asking for the extreme option ("strongly agree") produced 8% false positives, and an item asking for a non-extreme option 2% (Curran & Hauser, 2019). Announcing the check openly as an "attention check", for its part, may increase purposeful non-compliance (Lundmark et al., 2023). A single failed check is a noisy signal.

Acquiescence bias. Acquiescence is a serious problem, especially in adolescents: in data from ages 10–20 it varies widely among young people and strongly affects the psychometric properties of scales (Soto et al., 2008). Equal numbers of true-keyed and reverse-keyed items allow acquiescence to be controlled by within-person centring; correction for acquiescence has been shown in 18 countries to improve the congruence of factor loadings (Rammstedt et al., 2013). This explains why the keying imbalance of the IPIP-NEO-120 (Section 2.2) is not merely a technical detail: in a facet keyed in a single direction, an adolescent who "says yes" may appear high or low irrespective of his or her trait.

Intervention: warn or block? A real-time, gentle speeding warning reduced fast responses on subsequent questions in six web survey experiments (Conrad et al., 2017). A more forceful active warning (having the warning typed out and confirmed) slightly improved some indicators in a preregistered randomised experiment (N = 812, 300 IPIP-NEO items) (d = 0.18–0.25), made no difference to page time, longstring or attention-check items, and significantly increased drop-out; the authors noted that the improvement may be due in part to careless people leaving (Brühlmann et al., 2024). No study could be found that tested a practice of withholding the result until consistency is achieved. In clinical inventories, inconsistency scales derived from item pairs with similar or opposite content deem the profile "uninterpretable" at very high values; this is a narrowly defined exception logic.

Social desirability. The case for adding a separate "lie scale" is weak: social desirability scales cannot measure impression management and self-deception separately, and their construct validity for measuring faked responding is disputed. The ITC, too, recommends a proctored verification test after unproctored testing only for moderate- and high-stakes assessments; low-stakes use, in which the person obtains the information for his or her own interest, is defined separately (ITC, 2005).

Inference for this study. In a test taken before sign-up, with an audience that includes adolescents, careless responding is unavoidable, and making a hard decision on a single indicator also penalises those who respond attentively. For this reason we adopted a package of four indicators (page time, longstring, content pairs placed by design and an instructed-response item) and a graduated flow that comes into play when two or more indicators coincide (in the current form there is one instructed-response item in stage 1 and two in stage 2): the result is not blocked, an optional review is offered and, if necessary, a confidence note is added to the report; only in narrowly defined severe cases is an "uninterpretable" screen shown in place of the narrative report (ADR-026). Blocking was rejected because it increases drop-out and rewards appearing consistent. Attention-check items should ask for a non-extreme option and should not give the feeling of a hidden trap. Acquiescence bias, for its part, should be monitored with a separate index (this index today enters the same count as a fifth indicator; `Pusula-API/app/Scoring/V4/Quality.php`), and as a lasting solution the keying balance should be corrected in the single-direction facets (ADR-030). Because the thresholds in the literature rest on secondary accounts and have not been calibrated on Turkish or mobile data, they need to be re-tuned with pilot data (Chapter 4).

2.10 Abandonment and duration in tests taken before sign-up

A test taken before sign-up asks for effort at a moment when the participant has made no commitment whatsoever; for this reason the relationship between duration and abandonment is the primary constraint on the design. The common finding of the literature is that most of the attrition occurs at the start. In six online surveys with 1,963 undergraduates, about 10% of participants left immediately after the consent form, to which was added about 2 points of further drop-out for every 100 items (Hoerger, 2010). In a study in which the announced duration was varied experimentally, as the duration rose from 10 minutes to 30 minutes the numbers both starting and finishing fell, and the drop-out rate rose from 9.4% to 11.3%; however, the difference disappeared between the 20-minute and 30-minute versions. The same study showed that towards the end of the survey responses were faster,

shorter and less differentiated (Galesic & Bosnjak, 2009). The duration participants consider ideal is 10–15 minutes, and the duration they consider the maximum is 20–25 minutes; those who enjoy the survey accept longer durations (Revilla & Höhne, 2020).

The difference between devices is decisive. In mobile web surveys drop-out is between 2.6 and 12 times higher than on desktop (Mavletova & Couper, 2014); in a meta-analysis of mobile surveys the odds of drop-out, relative to a five-minute survey, increased by a factor of 1.09 at ten minutes and 1.42 at thirty minutes (Mavletova & Couper, 2015). On mobile, Likert grids increase straightlining; this effect is reduced when the items are presented each on its own row, in card form. There is also evidence for interface decisions. In a meta-analysis of 32 randomised experiments, a constant-speed progress bar was found to have no significant overall effect on drop-out, indicators that advance fast at first and then slowly reduced drop-out, and those that advance slowly at first increased it (Villar et al., 2013). Transition screens that contain no question trigger drop-out, because on these screens the participant reconsiders whether to continue.

Whether a long test increases perceived value is a separate question. In behavioural economics there are strong findings that effort increases perceived value, but they are conditional. The IKEA effect emerges only when the task is completed; those who leave it half finished assign a lower value (Norton et al., 2012). A service that shows the work being done in the background is perceived as more valuable even when the result is the same; however, when an unsuitable result is shown, displaying the effort increases dissatisfaction (Buell & Norton, 2011). Perceived accuracy, for its part, is not a good indicator of length, because people also find the generic feedback of short tests highly accurate (Dickson & Kelly, 1985; Poškus et al., 2014). Finally, where contact information is requested affects drop-out: in the U.S. Census Bureau's online survey test, drop-out almost doubled on the screen asking for contact information (0.9% versus 0.5%), and these screens were redesigned (Zelenak, 2016).

Inference for this study. The stage before sign-up was kept at about 12–13 minutes: within the ideal range on desktop, not far above the mobile threshold, and slightly longer than 16Personalities and Truity (ADR-026). The defensible reason for lengthening is not perceived accuracy but psychometrics; for this reason depth was left until after the user has invested effort, to a deepening stage after sign-up, and was made divisible into two sessions for those who wish. The drop-out literature supports three interface decisions: placing the psychometrically most critical blocks at the start and interspersing the scale items, showing progress by section and minimising transition screens. The e-mail step was first moved from the beginning of the test to the end of the last section, immediately before the open summary (ADR-026). Since 15 November 2026 there has been no separate e-mail step and no summary without sign-up: the test still starts anonymously, the account step comes when the sections are finished, scoring is done after the account has been created and the result is shown on the dashboard (ADR-073). Contact information is still requested not at the beginning of the test but at the end of the point at which the effort has been given.

2.11 Products on the market

Products in the consumer market fall roughly into two clusters: type or dimension tests of 60–130 items and 5–20 minutes (16Personalities, Truity, 123test, HIGH5) and in-depth instruments of 25–40 minutes (CliftonStrengths, PrinciplesYou, RHETI, CareerExplorer). Enterprise products (Hogan,

Pymetrics, Plum, TestGorilla) operate by employer invitation and are outside the scope of this work. Table 2.5 summarises the products comparable to a consumer funnel that precedes sign-up.

Table 2.5. Structure, free scope and psychometric transparency of competing consumer products.

Product	Model	Items and duration	Free scope	Paid scope	Published evidence
16Personalities	Five scales (NERIS), presented with MBTI letters	60 items, 7-point scale; about 10 min	Type and overview	Full report ₺179 (TR, review of 23 October 2026); Career Suite €29, Pro Suite €99	On its own site α .75–.87, retest .74–.83
Truity TypeFinder	4 dimensions, 23 facets (MBTI derivative)	130 items, 10–15 min	Short summary, without sign-up	Full report; no official price found	No figures on the page
Truity Career Personality Profiler	RIASEC and Big Five	94 items, 10–15 min	Short summary and suggestions	Full PDF report	No figures
CliftonStrengths	34 themes	177 items, about 30 min	None	Top 5 themes ≈ $20–25; 34 themes $59.99	Technical report; details could not be read
HIGH5	20 strengths	120 items, about 15 min	Top 5 and short description	Report $29	Not found
PrinciplesYou	17 dimensions, 28 archetypes	About 30 min	Entirely free	Team product	No figures
RHETI	Enneagram, 9 types	144 pairs, about 40 min	None	$20	Unsourced claim of "independent validity"
Turkey B2C (Kariyer Planım, Bizzr, Remetesting, CharacterIX)	Career and occupation tests	Mostly not stated	Short test (in some)	₺289–₺990; with ÖSYM matching ₺3,500	Not found in any

Note. Information is compiled from product pages dated 22 October 2026 and the research note (`Rehnüma kişilik testi psikometri/rakip_urunler.md`) and, for 16Personalities, from the authors' direct review of 23 October 2026. The Turkish lira price is the price seen in the review conducted from Turkey; that the difference from the euro prices stems from regional pricing is an inference.

16Personalities. Because it is the most visible product on the market, it was examined in detail. The company's own theory page states that it does not use Jung's cognitive functions, on the grounds that they are "very difficult to measure and validate scientifically", that it has "reworked and rebalanced" the Big Five dimensions, and that it uses MBTI-style letter abbreviations with an additional letter to accommodate five scales (16Personalities, n.d.-a). In the review the authors carried out by taking the test on 23 October 2026, the test consisted of 60 items and a seven-point response scale. The result was presented on five scales: the first four with MBTI letters (Mind I/E, Energy S/N, Nature T/F, Tactics J/P), and the fifth was the Identity (A/T) scale. The poles of the Identity scale, "self-assured and resistant to stress" and "sensitive to stress", correspond in essence to emotional stability, the reverse pole of Neuroticism; that is, the product presents the five dimensions of the Big Five in the language of the MBTI (inference; Table 2.4). A free report was provided and the full report was sold at a price of ₺179. When the text shown in blurred form in the locked sections of the full report was deciphered and examined by us, it turned out to be scrambled filler text; that is, what lay behind the lock was not real content but fake text giving the impression that content existed. On its own page the

company publishes, for the English version, α .75–.87 in a sample of 10,000 people and 5–7-month retest reliabilities of .74–.83 in 2,900 people; how the sample was collected is not explained (16Personalities, n.d.-c). The product is not regarded as a scientific instrument in mainstream personality psychology, and user reviews include the complaint that the paid report offers generic content not personalised to the responses (Trustpilot, n.d.).

Other products. At Truity no sign-up is required for the test and the short result; the account is created after the test, the result is not sent by e-mail and the user is drawn back to the site (Truity, n.d.). CliftonStrengths begins with payment and an access code; when the test is retaken both results are kept in the account and the user can select the active profile (Gallup, n.d.). PrinciplesYou shows the person the three archetypes, out of 28, that he or she most resembles; this form of presentation is close to a "soft" reading of the Gerlach clusters, but no independent peer-reviewed validation of the product could be found. No psychometric documentation was seen in any of the consumer-facing products in Turkey. As of 22 October 2026 no publicly available source could be found on a domestic product that the user took as a reference (Egonomist); its structure (Enneagram percentages, four-letter code, temperament, 9 strengths, 9 development areas, 45 competencies) is known only from the report in the user's possession (ADR-028). Deriving 45 competencies from five domain scores is a relabelling of the same variance and produces no new information (inference).

Inference for this study. The market leaves three gaps. The first is transparency: very few consumer products publish reliability figures, and in Turkey none do; publishing the Turkish α, retest and N values as they are collected is a differentiator in its own right. The second is honest preview: the fake text in the locked sections of 16Personalities shows how the tension between the free result and the paid report can erode trust. For this reason Rehnüma's trust page commits that no blurred or fake content will be shown in locked sections and that what will be unlocked will be stated by means of a table of contents (`Front-end/src/app/[locale]/(marketing)/analysis/trust/page.tsx`). On 22 November 2026 the presentation of the locked section changed: the section opens, the person reads a real opening sentence from his or her own result, and the remainder is blurred. The blurred part is not the person's real text but filler built from the introductory sentences of the other sections; the remainder of the text is not sent to the client (ADR-081). The sentence on the trust page has not yet been brought into line with this decision. Stating the free scope openly before the test also answers the common source of complaints about competitors. The third is context: no competitor could be found that links the report to daily tracking, goals and habits. There are also patterns worth copying: showing the full profile on all dimensions, presenting the several nearest archetypes by similarity, and keeping the two results side by side on retest.

2.12 Chapter summary

Inference for this study. The literature draws a coherent framework for the design of the test. In personality measurement the Five-Factor Model and the facet level provide the ground needed for a person-specific report; licence constraints reduce this ground, in a commercial Turkish product, to the IPIP-NEO-120. In interest measurement the single O*NET pool with its two lengths fits naturally into a two-stage form, but the fragility of the primary code and the licence's validation requirement limit what the report can claim. Work values and chronotype, with their low reliabilities or unclear licences, can be reported only in cautious language and as secondary layers. Type models are renamed

and dichotomised forms of the Big Five; the cut is the structural source of label instability at retest. The literature on response quality and drop-out supports a flow that warns rather than blocks and that starts with a short core and leaves depth for later. The market, for its part, leaves a gap in transparency, honest preview and connection to everyday life. The absence of a Turkish adaptation shows that no element of this framework is automatically valid for the Turkish version; how this evidence will be collected in stages is the subject of Chapter 4, and the details of the measurement design are the subject of Chapter 3.

Chapter references

16Personalities. (n.d.-a). *Our theory*. https://www.16personalities.com/articles/our-theory

16Personalities. (n.d.-c). *Reliability and validity*. https://www.16personalities.com/articles/reliability-and-validity

Adan, A., & Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: A reduced scale. *Personality and Individual Differences, 12*(3), 241–253. https://doi.org/10.1016/0191-8869(91)90110-W

Alwin, D. F., & Krosnick, J. A. (1985). The measurement of values in surveys: A comparison of ratings and rankings. *Public Opinion Quarterly, 49*(4), 535–552. https://doi.org/10.1086/268949

American Thoracic Society. (n.d.). *Munich ChronoType Questionnaire (MCTQ)*. https://site.thoracic.org/assemblies/srn/sleep-related-questionnaires/mctq

Atak, H. (2013). On Maddeli Kişilik Ölçeği'nin Türk kültürüne uyarlanması. *Nöropsikiyatri Arşivi, 50*(4), 312–319. https://doi.org/10.4274/npa.y6128

Atlı, A., & Keldal, G. (2017). [Vocational Personality Types Inventory: validity and reliability in high-school students]. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 18*(1), 73–93. https://dergipark.org.tr/tr/pub/ogusbd/article/330743 (reference to be verified)

Briley, D. A., & Tucker-Drob, E. M. (2012). Broad bandwidth or high fidelity? Evidence from the structure of genetic and environmental effects on the facets of the Five Factor Model. *Behavior Genetics, 42*(5), 743–763. https://doi.org/10.1007/s10519-012-9548-8

Brühlmann, F., Memeti, Z., Aeschbach, L. F., Perrig, S. A. C., & Opwis, K. (2024). The effectiveness of warning statements in reducing careless responding in crowdsourced online surveys. *Behavior Research Methods, 56*(6), 5862–5875. https://doi.org/10.3758/s13428-023-02321-z

Buell, R. W., & Norton, M. I. (2011). The labor illusion: How operational transparency increases perceived value. *Management Science, 57*(9), 1564–1579. https://doi.org/10.1287/mnsc.1110.1376

Cemalcılar, Z., Baruh, L., Kezer, M., Soto, C. J., Sümer, N., & John, O. P. (2021). Testing the BFI-2 in a non-WEIRD community sample. *Personality and Individual Differences, 182*, 111087. https://doi.org/10.1016/j.paid.2021.111087

Chauliac, M., Willems, J., Gijbels, D., & Donche, V. (2023). The prevalence of careless response behaviour and its consequences on data quality in self-report questionnaires on student learning. *Frontiers in Education, 8*, 1197324. https://doi.org/10.3389/feduc.2023.1197324

Chu, C., Russell, M. T., Hoff, K. A., Phan, W. M. J., & Rounds, J. (2022). What do interest inventories measure? The convergence and content validity of four RIASEC inventories. *Journal of Career Assessment, 30*(4), 776–801. https://doi.org/10.1177/10690727221081554

Conrad, F. G., Tourangeau, R., Couper, M. P., & Zhang, C. (2017). Reducing speeding in web surveys by providing immediate feedback. *Survey Research Methods, 11*(1), 45–61. https://ojs.ub.uni-konstanz.de/srm/article/view/6304 (reference to be verified)

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology, 102*(4), 874–888. https://neoacademic.com/2012/03/21/dont-use-abbreviated-personality-measures/

Curran, P. G., & Hauser, K. A. (2019). I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items. *Journal of Research in Personality, 82*, 103849. https://www.gwern.net/doc/psychology/personality/2019-curran.pdf

Dickson, D. H., & Kelly, I. W. (1985). The "Barnum effect" in personality assessment: A review of the literature. *Psychological Reports, 57*(2), 367–382. https://doi.org/10.2466/pr0.1985.57.2.367

Diener, E. (n.d.). *Satisfaction with Life Scale*. https://eddiener.com/scales/7

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*(2), 192–203. https://pubmed.ncbi.nlm.nih.gov/16768595/

Druiven, S. J. M., et al. (2020). Stability of chronotype over a 7-year follow-up period and its association with severity of depressive and anxiety symptoms. *Depression and Anxiety, 37*(5), 466–474. https://doi.org/10.1002/da.22995

Duckworth, A. (n.d.). *Research*. https://angeladuckworth.com/research/

EFPA. (2025). *EFPA review model for the description and evaluation of psychological and educational tests* (Version 2025.5). European Federation of Psychologists' Associations. https://www.infocop.es/wp-content/uploads/2025/09/efpa_test_review_model_v2025_5.pdf

Einarsdóttir, S., & Rounds, J. (2020). A quantitative review of gender differences in vocational interests in Iceland: Pervasive and persistent. *Nordic Journal of Transitions, Careers and Guidance, 1*(1), 10–26. https://doi.org/10.16993/njtcg.29

European Values Study. (n.d.). *The TRAPD method for survey translation*. https://europeanvaluesstudy.eu/surveys/surveys-evs/evs-2017/methodology/the-trapd-method-for-survey-translation/

Fischer, D., Lombardi, D. A., Marucci-Wellman, H., & Roenneberg, T. (2017). Chronotypes in the US: Influence of age and sex. *PLoS ONE, 12*(6), e0178782. https://doi.org/10.1371/journal.pone.0178782

Fuller, B. E., Holland, J. L., & Johnston, J. A. (1999). The relation of profile elevation in the Self-Directed Search to personality variables. *Journal of Career Assessment, 7*(2), 111–123. https://doi.org/10.1177/106907279900700202

Furnham, A., Moutafi, J., & Crump, J. (2003). The relationship between the revised NEO-Personality Inventory and the Myers-Briggs Type Indicator. *Social Behavior and Personality, 31*(6), 577–584. https://doi.org/10.2224/sbp.2003.31.6.577

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*(2), 349–360. https://academic.oup.com/poq/article-abstract/73/2/349/1832998

Gallup. (n.d.). *CliftonStrengths (StrengthsFinder)*. https://www.gallup.com/cliftonstrengths/en/strengthsfinder.aspx

Gerlach, M., Farb, B., Revelle, W., & Amaral, L. A. N. (2018). A robust data-driven approach identifies four personality types across four large data sets. *Nature Human Behaviour, 2*, 735–742. https://doi.org/10.1038/s41562-018-0419-z

Gerlach, M., Revelle, W., & Amaral, L. A. N. (2019). Reply to: Four personality types may be neither robust nor exhaustive. *Nature Human Behaviour, 3*(10), 1047–1048. https://doi.org/10.1038/s41562-019-0722-3

Glavin, K. W., & Savickas, M. L. (2011). Interpreting Self-Directed Search profiles: Validity of the "Rule of Eight". *Journal of Vocational Behavior, 79*(2), 414–418. https://doi.org/10.1016/j.jvb.2011.01.005

Goldberg, L. R., & Somer, O. (2000). The hierarchical structure of common Turkish person-descriptive adjectives. *European Journal of Personality, 14*(6), 497–531. https://doi.org/10.1002/1099-0984(200011/12)14:6%3C497::AID-PER379%3E3.0.CO;2-R

HEXACO. (n.d.). *The HEXACO Personality Inventory – Revised*. https://hexaco.org/hexaco-inventory

Hoerger, M. (2010). Participant dropout as a function of survey length in Internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior, and Social Networking, 13*(6), 697–700. https://doi.org/10.1089/cyber.2009.0445

Hoff, K. A., Briley, D. A., Wee, C. J. M., & Rounds, J. (2018). Normative changes in interests from adolescence to adulthood: A meta-analysis of longitudinal studies. *Psychological Bulletin, 144*(4), 426–451. https://www.pubmed.ncbi.nlm.nih.gov/29494193/

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4*(2), 97–110. https://pubmed.ncbi.nlm.nih.gov/1027738/

Hřebíčková, M., Jelínek, M., Květon, P., Benkovič, A., Botek, M., Sudzina, F., Soto, C. J., & John, O. P. (2020). Big Five Inventory 2 (BFI-2): Hierarchický model s 15 subškálami. *Československá psychologie, 64*(4), 437–460. https://webarchiv.lib.cas.cz/wayback/20201006133032/http:/cspsych.psu.cas.cz/result.php?id=1123

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

International Test Commission. (2005). *International guidelines on computer-based and internet-delivered testing*. https://www.intestcom.org/files/guideline_computer_based_testing.pdf

International Test Commission. (2017). The ITC guidelines for translating and adapting tests (2nd ed.). *International Journal of Testing, 18*(2), 101–134. https://doi.org/10.1080/15305058.2017.1398166

IPIP. (n.d.-a). *Permission*. International Personality Item Pool. https://ipip.ori.org/newPermission.htm

IPIP. (n.d.-b). *Translations*. International Personality Item Pool. https://ipip.ori.org/newTranslations.htm

IPIP. (n.d.-c). *Key to the NEO facets*. International Personality Item Pool. https://ipip.ori.org/newNEO_FacetsKey.htm

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Kajonius, P. J., & Johnson, J. A. (2019). Assessing the structure of the Five Factor Model of Personality (IPIP-NEO-120) in the public domain. *Europe's Journal of Psychology, 15*(2), 260–275. https://pmc.ncbi.nlm.nih.gov/articles/PMC7871748/

Katahira, K., Kunisato, Y., Yamashita, Y., & Suzuki, S. (2020). Commentary: A robust data-driven approach identifies four personality types across four large data sets. *Frontiers in Big Data, 3*, 8. https://doi.org/10.3389/fdata.2020.00008

Kelly, K. R., & Jugovic, H. (2001). Concurrent validity of the online version of the Keirsey Temperament Sorter II. *Journal of Career Assessment, 9*(1), 49–59. https://doi.org/10.1177/106907270100900104

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit. *Personnel Psychology, 58*(2), 281–342. https://doi.org/10.1111/j.1744-6570.2005.00672.x

Kung, F. Y. H., Kwok, N., & Brown, D. J. (2018). Are attention check questions a threat to scale validity? *Applied Psychology, 67*(2), 264–283. https://doi.org/10.1111/apps.12108

Lipnevich, A. A., Credé, M., Hahn, E., Spinath, F. M., Roberts, R. D., & Preckel, F. (2017). How distinctive are morningness and eveningness from the Big Five factors of personality? A meta-analytic investigation. *Journal of Personality and Social Psychology, 112*(3), 491–509. https://www.ncbi.nlm.nih.gov/pubmed/27977220

Liu, J., Hoff, K. A., Chu, C., Oswald, F. L., & Rounds, J. (2025). Toward whole-person fit assessment: Integrating interests, values, skills, knowledge, and personality using O*NET. *Journal of Applied Psychology, 110*(5), 623–647. https://doi.org/10.1037/apl0001232

Lundmark, S., Silber, H., Rossmann, J., & Gummer, T. (2023). *Comparing types of attentiveness measures*. University of Gothenburg. https://www.gu.se/sites/default/files/2023-09/Comparing%20types%20of%20attentiveness%20measures%20-%20Lundmark%20Silber%20Rossman%20%20Gummer%20%282023%29.pdf (reference to be verified)

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19–40. https://psychology.psy.sunysb.edu/attachment/measures/content/maccallum_on_dichotomizing.pdf

Mavletova, A., & Couper, M. P. (2014). [Breakoff rates in mobile and desktop web surveys; GOR 2014 conference presentation]. https://cts.hse.ru/data/2014/04/03/1317876257/Mavletova_Couper_GOR_2014.pdf (reference to be verified)

Mavletova, A., & Couper, M. P. (2015). A meta-analysis of breakoff rates in mobile web surveys. In D. Toninelli, R. Pinter, & P. de Pedraza (Eds.), *Mobile research methods*. Ubiquity Press. https://dx.doi.org/10.5334/bar.f (reference to be verified)

McCloy, R., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999a). *Development of the O*NET paper-and-pencil Work Importance Locator*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/DevWIL.pdf

McCloy, R., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999b). *Development of the O*NET computerized Work Importance Profiler*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/DevCWIP.pdf

McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality, 57*(1), 17–40. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Morillo, D., Abad, F. J., Kreitchmann, R. S., Leenen, I., Hontangas, P., & Ponsoda, V. (2019). The journey from Likert to forced-choice questionnaires: Evidence of the invariance of item parameters. *Revista de Psicología del Trabajo y de las Organizaciones, 35*(2), 75–83. https://doi.org/10.5093/jwop2019a11

Morsünbül, Ü. (2014). [Turkish adaptation of the Quick Big Five Personality Test: validity and reliability]. *Düşünen Adam, 27*, 316–322. https://dusunenadamdergisi.org/storage/upload/pdfs/1585750117-en.pdf (reference to be verified)

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology, 22*(3), 453–460. https://doi.org/10.1016/j.jcps.2011.08.002

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science, 7*(4), 384–403. https://europepmc.org/abstract/MED/26168474

O*NET Resource Center. (n.d.-a). *O*NET Career Exploration Tools license*. https://www.onetcenter.org/license_tools.html

O*NET Resource Center. (n.d.-b). *O*NET Tools Developer License*. https://www.onetcenter.org/license_toolsdev.html

O*NET Resource Center. (n.d.-c). *O*NET Work Importance Profiler and Work Importance Locator*. https://www.onetcenter.org/WIP.html

O*NET Resource Center. (n.d.-d). *O*NET Database license*. https://www.onetcenter.org/license_db.html

PAR. (n.d.). *Self-Directed Search handbook, Chapter 1*. Psychological Assessment Resources. https://www.parinc.com/Portals/0/Webuploads/samplerpts/SDS%20Handbook%20Chapter%201.pdf

Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*(3), 524–539. https://doi.org/10.1037/0022-3514.81.3.524

Perkmen, S., & Tezci, E. (2015). [A 55-item Holland-based interest scale in pre-service teachers]. Balıkesir University open access. https://acikerisim.balikesir.edu.tr/items/d7bef6e3-02d7-4e88-91b1-38e45092a012/full (reference to be verified)

Pittenger, D. J. (1993). The utility of the Myers-Briggs Type Indicator. *Review of Educational Research, 63*(4), 467–488. https://doi.org/10.3102/00346543063004467

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research, 57*(3), 210–221. https://doi.org/10.1037/1065-9293.57.3.210

Poškus, M. S., Kairys, A., Liniauskaitė, A., & Žukauskienė, R. (2014). "Your NEO summary": Is it a valid way of providing feedback, or is its accuracy due to the Barnum effect? *Psichologija, 49*, 44–59. https://doi.org/10.15388/Psichol.2014.49.3694

Pündük, Z., Gür, H., & Ercan, İ. (2005). Sabahçıl-akşamcıl anketi Türkçe uyarlamasında güvenilirlik çalışması. *Türk Psikiyatri Dergisi, 16*(1), 40–45. https://avesis.uludag.edu.tr/yayin/99c6a668-bc64-4417-bd78-6a3a55b1d31c/a-reliability-study-of-the-turkish-version-of-the-mornings-evenings-questionnaire-sabahcil-aksamcil-anketi-turkce-uyarlamasinda-guvenilirlik-calismasi

Purdue University. (n.d.). [The power of expressed and measured interests to predict current choice; doctoral dissertation]. https://docs.lib.purdue.edu/dissertations/AAI8400378 (reference to be verified)

Rammstedt, B., Kemper, C. J., & Borg, I. (2013). Correcting Big Five personality measurements for acquiescence: An 18-country cross-cultural study. *European Journal of Personality, 27*(1), 71–81. https://doi.org/10.1002/per.1894

Revilla, M., & Höhne, J. K. (2020). How long do respondents think online surveys should be? New evidence from two online panels in Germany. *International Journal of Market Research, 62*(5), 538–545. https://www.sowi.uni-mannheim.de/media/Einrichtungen/gip/Bilder/Dokumente/revilla_hoehne__2020__-_online_survey_length.pdf

Roenneberg, T., Pilz, L. K., Zerbini, G., & Winnebeck, E. C. (2019). Chronotype and social jetlag: A (self-) critical review. *Biology, 8*(3), 54. https://europepmc.org/article/PMC/PMC6784249

rOpenSci. (n.d.). *mctq: Tools to process the Munich ChronoType Questionnaire*. https://packages.ropensci.org/mctq/doc/manual.html

Rounds, J., Wee, C. J. M., Cao, M., Song, C., & Lewis, P. (2016). *Development of an O*NET Mini Interest Profiler (Mini-IP) for mobile devices: Psychometric characteristics*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/Mini-IP.pdf

Rounds, J., & Su, R. (2014). The nature and power of interests. *Current Directions in Psychological Science, 23*(2), 98–103. https://doi.org/10.1177/0963721414522812

Rounds, J., Su, R., Lewis, P., & Rivkin, D. (2010). *O*NET Interest Profiler Short Form psychometric characteristics: Summary*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/IPSF_Psychometric.pdf

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology, 38*(2), 173–212. https://www.um.edu.mt/library/oar/bitstream/123456789/

21679/1/The%20geographic%20distribution%20of%20Big%20Five%20personality%20traits%20Patterns%20and%20profiles%20of%20human%20self-description%20across%2056%20nations.pdf

Somer, O., Korkmaz, M., & Tatar, A. (2002). Beş Faktör Kişilik Envanteri'nin geliştirilmesi-I: Ölçek ve alt ölçeklerin oluşturulması. *Türk Psikoloji Dergisi, 17*(50), 21–36. https://avesis.mcbu.edu.tr/yayin/9f580945-7e2e-48e8-9a28-d68cfc6c173e/development-of-five-factor-personality-inventory-bes-faktor-kisilik-envanterinin-gelistirilmesi-i-olcek-ve-alt-olceklerin-olusturulmasi

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143. https://pubmed.ncbi.nlm.nih.gov/27055049/

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of Big Five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718–737. https://doi.org/10.1037/0022-3514.94.4.718

Stein, R., & Swan, A. B. (2019). Evaluating the validity of Myers-Briggs Type Indicator theory: A teaching tool and window into intuitive psychology. *Social and Personality Psychology Compass, 13*(2), e12434. https://doi.org/10.1111/spc3.12434

Stoll, G., Rieger, S., Lüdtke, O., Nagengast, B., Trautwein, U., & Roberts, B. W. (2017). Vocational interests assessed at the end of high school predict life outcomes assessed 10 years later over and above IQ and Big Five personality traits. *Journal of Personality and Social Psychology, 113*(1), 167–184. https://europepmc.org/search?query=TITLE%3A%22Vocational%20interests%20assessed%20at%20the%20end%20of%20high%20school%22

Süerdem, A., & Erkök, B. (2016). Assessing the reliability and validity of a shorter version of RIASEC in Turkish. *SHS Web of Conferences, 26*, 01063. https://doi.org/10.1051/shsconf/20162601063

Tatar, A. (2017). [Turkish translation of the Big Five-50 Personality Test and its comparison with the short form of the Five Factor Personality Inventory]. *Anadolu Psikiyatri Dergisi, 18*(1), 51–61. https://search.trdizin.gov.tr/yayin/detay/266100 (reference to be verified)

Tatar, A. (2018). [Psychometric properties of the Turkish long and short forms of the HEXACO-PI-R]. *Anadolu Psikiyatri Dergisi, 19*(Suppl. 1). https://www.imrpress.com/journal/AP/19/Supplement%201/10.5455/apd.283098 (reference to be verified)

The Myers-Briggs Company. (n.d.). *Trademarks*. https://themyersbriggs.com/en-US/Support/Trademarks

Truity. (n.d.). *TypeFinder personality test*. https://www.truity.com/test/type-finder-personality-test-new

Trustpilot. (n.d.). *16Personalities reviews*. https://ca.trustpilot.com/review/16personalities.com

Tsaousis, I. (2010). Circadian preferences and personality traits: A meta-analysis. *European Journal of Personality, 24*(4), 356–373. https://doi.org/10.1002/per.754

Twomey, C., & Johnson, J. A. (2022). Open-source personality trait norms for the United Kingdom and Ireland. *European Journal of Psychological Assessment, 38*(6), 432–439. https://doi.org/10.1027/1015-5759/a000644

Uygur, H., Tekdemir, R., Uygur, Ö. F., Aydın, E. F., Çelik, M., Babacan, H. E., et al. (2024). Psychometric properties of the Turkish reduced Morningness-Eveningness Questionnaire. *Chronobiology International, 41*(5), 632–646. https://doi.org/10.1080/07420528.2024.2339964

Vardarlı, B., Özyürek, R., Wilkins-Yel, K. G., & Tracey, T. J. G. (2017). Examining the structure of vocational interests in Turkey in the context of the personal globe model. *International Journal for Educational and Vocational Guidance, 17*(3), 347–359. https://doi.org/10.1007/s10775-016-9338-6

Vedel, A., Wellnitz, K. B., Ludeke, S., Soto, C. J., John, O. P., & Andersen, S. C. (2021). Development and validation of the Danish Big Five Inventory-2: Domain- and facet-level structure, construct validity, and reliability. *European Journal of Psychological Assessment, 37*(1), 42–51. https://doi.org/10.1027/1015-5759/a000570

VIA Institute on Character. (n.d.). *Terms of service*. https://www.viacharacter.org/terms-of-service

Villar, A., Callegaro, M., & Yang, Y. (2013). Where am I? A meta-analysis of experiments on the effects of progress indicators for web surveys. *Social Science Computer Review, 31*(6), 744–762. https://openaccess.city.ac.uk/id/eprint/14427/

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology, 74*, 577–596. https://papers.ssrn.com/abstract=4337272

Wasti, S. A., Lee, K., Ashton, M. C., & Somer, O. (2008). Six Turkish personality factors and the HEXACO model of personality structure. *Journal of Cross-Cultural Psychology, 39*(6), 665–684. https://doi.org/10.1177/0022022108323783

Welz, M., & Alfons, A. (2023). *When respondents don't care anymore: Identifying the onset of careless responding* [Preprint]. arXiv. https://arxiv.org/abs/2303.07167

Zelenak, M. F. (2016). *2014 American Community Survey (ACS) Internet test* (American Community Survey Research and Evaluation Report Memorandum Series No. ACS16-RER-03). U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2016/acs/2016_Zelenak_01.pdf

3. Measurement Design

This chapter describes the measurement architecture of the Rehnüma character analysis: the principles by which the test was divided into two stages, how many items are asked from which item pool at each stage, how the items were rendered into Turkish, which indicators monitor response quality, how raw answers are turned into scores and scores into the readable layers of the report (mind motif, temperament, archetype, natural tendency, strength), and finally how occupation matching and personality percentiles against the published norm are computed. The account rests on the decision records (ADR-026, ADR-028, ADR-029, ADR-030, ADR-032, ADR-044, ADR-082) and on the research reports; however, every number, threshold and formula has been verified by re-reading it from the source code and data files of the test service. Where the code and a document differ, the code has been taken as authoritative and the difference is indicated in a "*Note*." line.

The scope of the chapter is the *design* of the measurement; the evidence on how well the design works (simulation findings and the pilot plan) is treated in Chapter 4, and the way the report is presented to the user in Chapter 5. The structure described here corresponds to the scoring version whose `scoring_version` value on the service side is `5.0` (`Pusula-API/config/pusula.php`). The current form is form v5 (ADR-044): it carries over the Likert core of form v4 and its expressed-interest, values and rhythm sections unchanged, and adds two ranking steps on top. The chapter therefore describes the v4 core first and then the steps added by v5; the Gate 0 simulation was run with form v4 (Chapter 4).

3.1 Design principles

Form v4 was rebuilt after the previous version (v3) was found to be "too short, too easy, with a simple report" (ADR-026). In v3 a full ranking of the six Holland areas was blended fifty-fifty with the Likert measurement, and the Big Five was measured with 25 items. The blend had no published precedent, and without going down to facet level the strengths narrative could not get beyond a few sentences. The three principles of the new design came out of this diagnosis.

3.1.1 Preliminary value before sign-up

The dominant pattern in the consumer market is "test first, then a short free result, then a paid full report"; asking for contact details at the start of the flow increases drop-out (evidence grade: weak; research report, "Competitors and funnel"). Form v4 was therefore designed with a flow that starts anonymously and gives an interpretable summary before sign-up. This end of the flow changed later (ADR-073): the test still starts anonymously and the account step is still at the end of the sections, but the result is not shown without an account; scoring is carried out after the account has been created and the result appears on the dashboard. The names "pre-sign-up" and "post-sign-up" continue to be used in this chapter as the design names of the two stages. For the design, the consequence is this: the pre-sign-up part has to produce a meaningful measurement on its own. The five Big Five domains, the six RIASEC interest areas and the first occupation suggestions based on them must be com-

putable reliably at this stage. Facet scores, by contrast, can remain only at the level of a "first impression" at this stage, because two items per facet provide sufficient reliability for the domain score but not for the facet score (ADR-026, Rationale; Johnson, 2014).

3.1.2 Two stages

The second principle is the division of the measurement into two *resolutions*: a pre-sign-up core and a post-sign-up deepening. The critical design decision is that the second stage does not repeat the first but completes it. In personality, two of the four items of each facet are asked in the first stage and the remaining two in the second; in interests, the 30 items of the O*NET Mini Interest Profiler (Mini-IP) come in the first stage and the remaining 30 items of the Short Form (IP-SF) in the second. Because the Mini-IP items are a subset of the Short Form (Rounds et al., 2016), the sum of the two stages gives exactly the published 60-item form. The free summary and the full report thus become two resolutions of the same measurement, and no item is wasted. The finding of Gerlach and colleagues (2018) further supports this ordering: whereas four robust personality clusters separated in the 120- and 100-item forms, only two of them separated in the version reduced to 44 items; the pre-sign-up profile assignment should therefore be marked as a "preliminary" assignment.

In the code the two stages are modelled as two states of a single entity. When the form is created, all sections are generated and stored together with a `stage` field (1 or 2); on every call `AssessmentScorer::completeV4` sets the target `completed_stage + 1`, and validates and scores *all* sections up to that stage together (`Pusula-API/app/Assessment/AssessmentScorer.php`). When stage 2 is completed, the stage 1 answers are reused; the second scoring overwrites the first. The same function does not allow a stage with unanswered items to be completed (`requireComplete: true`).

3.1.3 Time budget

The third principle is the time budget. In the literature the perceived ideal length of an online survey has been reported as 10–15 minutes and the maximum length as 20–25 minutes (Revilla & Höhne, 2020); that marked drop-out begins after about 9 minutes on mobile is reported as vendor data (evidence grade: weak). Most drop-out occurs at the beginning, and thereafter the slope per item is moderate (as reported from Hoerger's data; research report, "Duration"). The "IKEA effect", which raises the value of a completed task, appears only when the task has been finished (Norton et al., 2012). Together these three findings yielded the following rule: the pre-sign-up core should not exceed 12–13 minutes, and depth should be left until after the person has invested their effort, that is, until after sign-up (ADR-026 §1).

The code produces an estimated duration for each section: in Likert sections `ceil(number of items / 8)` minutes, that is, an assumption of eight items per minute (`FormV4::personality`, `FormV4::interest`). This assumption is also flagged in the research report as "not verified against a source; to be measured with product data". Under this rule, in the form v4 core stage 1 comes to 13 minutes in total (4 + 4 + 4 + 1) and stage 2 to 18 minutes (8 + 4 + 5 + 1); the two stages together take about 31 minutes and match the "~~12–13 min + ~15–18 min" band in ADR-026. For the two ranking steps of form v5 the duration is hard-coded (`FormV5`: self-placement 1 min, statement ranking 10 min), with these, the `minutes` total in the code is 14 in stage 1, 28 in stage 2 and 42 minutes~~

~~together. ADR-044 refers to the second stage as "~~25 min" and retains the 12-minute claim for the first stage; the totals in the code are slightly above these two values. Stage 2 has been designed so that it can be split into two sessions if the person wishes.

In the interface, Likert items are presented six per page, each as a card on its own row (`Frontend/src/components/test/SectionRunner.tsx`, `PAGE_SIZE = 6`); on mobile, Likert grids increase straightlining (research report, "Duration"). Page times reach the server in the `elapsed` field and serve as input to the speed indicator (3.6).

Note. ADR-026 §1 states that the items are "interleaved across scales within the same session". In the code the interleaving is at section level: items are shuffled *within* a section (`FormV4::shuffle`), whereas the sections come in a fixed order (self-placement → personality A → interests → personality B → expressed interest; in stage 2, statement ranking → personality C → interest completion → values → rhythm). Personality and interest items are not presented mixed on the same page. Splitting the personality items into two blocks and placing the interest block between them achieves part of the fatigue-spreading effect intended by the ADR (inference).

3.2 Form structure: the v4 core and the two ranking steps of v5

The class that builds the form v4 core is `Pusula-API/app/Assessment/FormV4.php`. The form presented to the person today comes from `FormV5`: this class extends `FormV4`, takes its eight sections as they are, and adds the `self_place` section at the very beginning of stage 1 and the `rank` section at the very beginning of stage 2 (`Pusula-API/app/Assessment/FormV5.php`; ADR-044). `FormBuilder` calls only `FormV5` and produces the version that goes to the client. The structure of the core derives from the four data files from which the items are loaded into the database: `ipip_neo120.json` (120 personality items, 30 facet definitions, 2 attention-check items), `interest_items_onet.json` (60 interest items, 1 attention-check item), `work_values.json` (20 cards, 6 values) and `rhythm.json` (3 questions) (`Pusula-API/database/data/`). The content of the two ranking steps comes from the eight archetype cards in `report_texts.json` and from the 64 statements in the file `archetype_statements.json` (3.2.3). Table 3.1 gives the ten sections of the form, with the numbers counted from the data files and from the section constructors in the code.

Table 3.1. Sections of the form (the v4 core and the two ranking steps of v5), stage and item counts.

Stage	Section key	Section type	Content	Items	Attention-check item	Estimated duration (min)
1	`self_place`	`self_place`	Eight archetype cards: "closest 3 (ordered) + farthest 1" (v5)	8 cards	—	1
1	`personality_a`	`likert`	IPIP-NEO-120, 30 facets × 1 core item	30	—	4
1	`interest`	`likert`	O*NET Mini-IP (the 30 items of the IP-SF marked `mini`)	30	—	4
1	`personality_b`	`likert`	IPIP-NEO-120, 30 facets × 1 core item	30	1 (`check:1`, expected 2)	4
1	`expressed`	`expressed`	Six RIASEC areas: "most 2 + least 1"	6 options	—	1

Stage	Section key	Section type	Content	Items	Attention-check item	Estimated duration (min)
2	`rank`	`rank`	8 rounds × 8 statements (one from each archetype), full ranking (v5)	64 statements	—	10
2	`personality_c`	`likert`	The 60 non-core items of the IPIP-NEO-120	60	1 (`check:2`, expected 4)	8
2	`interest_more`	`likert`	The 30 items of the IP-SF outside the Mini-IP	30	1 (`check:3`, expected 2)	4
2	`values`	`card_sort`	O*NET Work Importance Locator cards, 5 columns × 4	20 cards	—	5
2	`rhythm`	`choice`	Rehnüma's three rhythm preference questions	3	—	1
	Stage 1 total			**90 Likert + 8 archetype cards + 1 selection screen**	**1**	**14**
	Stage 2 total			**90 Likert + 64 statements + 20 cards + 3 questions**	**2**	**28**

Note. Item counts were counted from the `Pusula-API/database/data/*.json` files and cross-checked against `Pusula-API/tests/Feature/FormV4Test.php` (`test_iki_asamali_form_gelir`): the test expects the form version to be 5, the section order to be as in the table and, including attention-check items, 31 items for `personality_b`, 61 for `personality_c` and 31 for `interest_more`; the counts of the two ranking steps are tested by `FormV5Test` (`test_form_v5_iki_siralama_adimi_tasir`). Likert durations are the `ceil(n/8)` rule in `FormV4` applied to the number of items including the attention-check item; in the other sections the duration is hard-coded (self-placement 1, statement ranking 10, card sort 5, expressed interest and rhythm 1 min). The eight sections other than the two rows marked "(v5)" are the form v4 core; when only the core is counted, the stage totals are 13 and 18 minutes. Across the two stages there are in total 180 content Likert items (120 personality, 60 interest), 3 attention-check items, 20 cards, 3 rhythm questions, one expressed-interest screen, eight archetype cards and 64 ranking statements.

Two structural features can be read from the table. First, the personality measurement is distributed over three blocks, and blocks A and B mirror each other: each facet is represented by one item in each of these two blocks. `FormV4::build` groups the core items by facet; it puts the first core item of each facet into A and the second into B. The test checks this with the assertion "the two core items of the same facet are split between A and B". The function of this split is to obtain two independent observations for each facet even in stage 1 and so to make odd–even consistency computable (3.6). Second, the first stage contains only a single attention-check item, and this item is embedded in the 31-item second personality block; the second stage contains two attention-check items.

Note. The stage table of ADR-026 states "1 instructed-response item" for stage 2; in the code there are two instructed-response items in stage 2 (`FormV4::CHECK_IDS`: `personality_b` → −1, `personality_c` → −2, `interest_more` → −3). The second check was listed as open work in the "Impact" section of ADR-029 ("a second attention check for stage 2") and was implemented in the code with the comment in `FormV4::interest` ("one check was too few for 60 items"). In this respect the open-work list of ADR-029 does not reflect the current state.

3.2.1 Section types

Six section types are used in the form: four come from the form v4 core and two (`self_place`, `rank`) from form v5. In addition to these, the validator (`Pusula-API/app/Assessment/AnswerValidator.php`) also recognises two types left over from v3 (`forced_choice`, `pick_top`), but the current form does not

produce them. Table 3.2 summarises the answer format and the validation rule of each type.

Table 3.2. Section types, answer formats and server-side validation rules.

Type	Sections	Answer format	Validation rule	Scorer
`likert`	`personality_a/b/c`, `interest`, `interest_more`	item id → integer 1–5	An item not in the form is rejected; the value cannot be outside `Likert::MIN`–`MAX`; on completion every item must have been answered	`PersonalityScorer`, `FacetScorer`, `InterestScorer`, `Quality`
`expressed`	`expressed`	`{most: [id, id], least: [id]}`	The two sets are disjoint; on completion exactly 2 + 1 selections	`ExpressedInterest`
`card_sort`	`values`	card id → column 1–5	The value must be one of the columns; at most 4 cards in a column; on completion all 20 cards distributed	`ValuesScorer`
`choice`	`rhythm`	question id → option key (a–e)	The key must be among the options of that question	`RhythmScorer`
`self_place`	`self_place`	`{most: [id, id, id], least: [id]}`	Same rule as `expressed`; on completion exactly 3 + 1 selections	`SelfPlacement`
`rank`	`rank`	round number → ordered statement ids	Each round must be a permutation of the eight statements of that round; on completion all eight rounds ranked	`RankScorer`

Note. The validator applies all rules on the server; the client cannot alter the result and can only answer the items in the form presented to it. Source: `Pusula-API/app/Assessment/AnswerValidator.php`.

The two Likert scales use different label sequences. Personality items are answered with "Not at all like me / Not much like me / Neither like nor unlike me / Quite like me / Exactly like me" (Turkish: "Bana hiç uymuyor / Bana pek uymuyor / Ne uyuyor ne uymuyor / Bana oldukça uyuyor / Bana tamamen uyuyor"); interest items with the sequence "Strongly dislike / Dislike / Unsure / Like / Strongly like" (Turkish: "Hiç hoşlanmam / Hoşlanmam / Emin değilim / Hoşlanırım / Çok hoşlanırım") (`FormV4::personality`, `FormV4::interest`). The personality instruction asks the person to think of "who you usually are, not who you would like to be", to compare themselves with people their age and to go with their first thought; the last sentence of the instruction is "A few items only check that you are paying attention." (ADR-026, Addendum). The interest instruction reminds the person that training, experience and pay do not matter; this is the counterpart of the O*NET Interest Profiler's instruction to "think only about whether you would like it".

3.2.2 Separating the keys from the client

The form is stored on the server together with its scoring keys (RIASEC code, facet, reverse-item marker, expected answer of the attention-check item, scores of the rhythm options); only the item text and the options go to the client, through `FormBuilder::toPublic`. The test checks this separation by verifying that none of the keys `"keyed"`, `"facet"`, `"expected"`, `"score"` and `"code"` is present in the JSON sent to the client. The only exception is the expressed-interest section: there each option

carries its RIASEC letter in the `code` field; the title is already the name of the area, and the letter is used only as a language-independent identifier (ADR-046). In the two ranking steps the archetype key does not go to the client; a card is presented only with its name and one-line description, and a statement only with its text and round number (`FormBuilder::toPublic`). This design has two consequences. In terms of security, the person cannot see which item belongs to which dimension or which item is reverse-scored; this makes "looking for the right answer" more difficult. In terms of research, the simulation script has to reach the hidden keys not through the API but by reading the local database directly (4.3).

The order of the items within a section is shuffled once, at the moment the form is created, and stored. This order is the basis of the "order of presentation" definition of the consecutive-identical-response (longstring) indicator (3.6).

3.2.3 The two ranking steps of form v5

In form v4 the person nowhere stated a holistic judgement about themselves; the report was read in one direction only. ADR-044 filled this gap with two within-person (ipsative) steps and did not touch the Likert bank (the decision of ADR-030 that "items do not change before the pilot" remains in force for the Likert items).

Self-placement (`self_place`) is the first section of stage 1. The person is shown eight archetype cards (name and one-line description; `report_texts.json` → `archetypes`) in shuffled order; the person selects, in order, the three they find closest to themselves and the one they find farthest. That the step comes before the items is deliberate: the self-view is taken without being influenced by having answered the measurement items (`FormV5` class comment).

Statement ranking (`rank`) is the first section of stage 2. In each of eight rounds, one first-person statement from each of the eight archetypes (such as "In a hard moment I stay calm and pull the room together.") arrives in shuffled order and the person fully ranks all eight; the 64 statements in total are in the file `archetype_statements.json`. The note in the file records that each round is balanced within itself for desirability and length and that all the statements are positive or neutral; this balance is a writing principle, it has not been measured, and testing it in the pilot is open work (ADR-044, Impact). The interface presents each round on a separate page (`Front-end/src/components/test/SectionRunner.tsx`).

The common feature of the two steps is that they do not enter the Five-Factor and interest scores: Likert scoring and occupation matching do not see these answers; the quality indicators, too, are computed only from the answers of the Likert sections (the only indirect contact is page time; 3.6.2). Their scoring is described in 3.7.5 and their juxtaposition with the measurement in the report (the three-way comparison) in 3.8.6. Within-person ranking does not produce a percentile or a norm; for this reason the steps were placed alongside the Likert items, not in their place (ADR-044, Rationale).

3.3 Item pools

The four item pools in the form v4 core of the test have different scientific standing, and the report has to preserve this difference: the Big Five layer stands on the firmest ground, the interest layer is strong at source but subject to licence conditions, the work values layer is weak even at source, and

the rhythm layer has no psychometric validation at all (validation protocol report, "The source instruments are strong"). The eight archetype cards and 64 ranking statements added by form v5 are Rehnüma's original content and, like the rhythm questions, have no psychometric validation (3.2.3). Table 3.3 sets side by side the source and licence of the four pools, their use in Rehnüma and the psychometric values reported by the source itself.

Table 3.3. Item pools: source, licence, use in Rehnüma and published psychometric values of the source instruments.

Pool	Source	Licence	In Rehnüma	Values reported at source	Turkish evidence
IPIP-NEO-120	Johnson (2014)	Public domain, commercial use permitted	120 items; 4 per facet, 2 in stage 1	Domain α .81–.90; facet α .63–.88 (N = 619,150); at domain level corrected r = .94 with the NEO PI-R	No published adaptation
O*NET IP-SF / Mini-IP	National Center for O*NET Development; Rounds et al. (2016)	O*NET Tools Developer License; validation study and notice in case of modification (§3(b))	60 items; the 30 of the Mini-IP in stage 1	IP-SF α .78–.87; retest .78–.86 (N = 132); circular order CI = .69; first-letter κ = .73 between Mini-IP and IP-SF	No published adaptation
O*NET Work Importance Locator	McCloy et al. (1999b); WIL instructions	O*NET Tools Developer License	20 cards, 6 values, 5 columns × 4	Value-level retest r = .59–.66; α .50–.86; top value ranked first again in 62–70%	No published adaptation
Rhythm questions	Rehnüma (original)	—	3 preference questions	No validation; the rMEQ was not used because its commercial licence is unclear	—

Note. The WIP-C values were taken from the development report of the computerised Work Importance Profiler (McCloy et al., 1999b); the reliability coefficients of the paper WIL itself were not read in the research notes (`turetilmis_katmanlar.md` §7, "Gaps").

3.3.1 IPIP-NEO-120

The personality pool is the 120-item form that Johnson (2014) shortened from the 300-item IPIP-NEO. The form was developed from an internet sample of 21,588 people and tested in samples of 481, 307,313 and 619,150 people. For each facet, the four items that kept alpha highest were selected from the ten-item pool by successive elimination; in item selection, equality of the numbers of positively and reverse-keyed items was not taken as a criterion, and alpha was prioritised (Johnson, 2014; the consequences of this decision are in 3.5). At facet level the weakest scales were reported as O6 Liberalism (.63), O3 Emotionality (.65) and C3 Dutifulness (.67); in a US sample of 320,128 people, O6, O3, C3 and A4 remained below .70, and the second-order loading of O6 on the Openness domain stayed at .35 (Kajonius & Johnson, 2019). Johnson's own judgement draws the boundary of the product language: the facet scales are adequate for research and should probably not be used for important decisions about individuals.

In the data file each domain has 24 items and each facet 4 (`ipip_neo120.json`). Two structural differences stand out. First, the neuroticism domain (N) is scored in the report under the name "emotional stability" (S) and in the reverse direction: the N facets take the codes S1–S6 (S1 Managing worry, S2 Managing anger, S3 Mood balance, S4 Social ease, S5 Impulse control, S6 Stress resilience) and their keys have been reversed. This is a product-language decision that does not change the measurement.

Second, the items of the O6 facet that rest on US politics (voting for liberal/conservative candidates, being tough on crime) have been rewritten, and the facet has been given the name "Flexibility in values"; in the file three items are marked with the prefix `[Rehnüma]` or `[Rehnüma uyarlaması]` ("Rehnüma adaptation") (3.4.2).

Two of the four items of each facet are marked `core: true` and are asked in stage 1; this makes 60 core items in total, 12 per domain. Indirect evidence for the adequacy of 12 items per domain for the domain score is the alpha values of .79–.87 of the IPIP-50 versions with 10 items per domain (research report, "Big Five"). The selection of core items aims to have each facet represented in stage 1 by one positively keyed and one reverse-keyed item; however, the keying structure of the source does not permit this in every facet (3.5).

Note. The comment line of the `FormV4` class states that the two core items of the same facet are "one positively and one reverse-keyed". In the data file this is true of only 21 facets; in nine facets (S1, E5, E6, O1, A2, A4, A5, C1, C6) all four items are keyed in the same direction, so that the two core items are also in the same direction. The domain-level keying distribution of the core set is given in Table 3.4.

3.3.2 O*NET Interest Profiler Short Form and Mini-IP

The interest pool is the 60 items of the O*NET Interest Profiler Short Form (IP-SF): ten activity statements for each of the six RIASEC areas (such as "Build kitchen cabinets" and "Buy and sell stocks and bonds"). The items of the IP-SF were selected from the 180-item long form using item means, distributions, cross-correlations and two-dimensional spatial maps; alpha is .78–.87 (mean .81) in the development sample, and retest r = .78–.86 in the stability sample of 132 people; in the randomisation test of the RIASEC circular order the correspondence index (CI) was found to be .69, and two-dimensional scaling accounted for 99% of the variance (National Center for O*NET Development, IPSF psychometric report). Gender differences are large (d = .86 in the Realistic area, d = −.59 in the Social area).

The Mini-IP is a 30-item subset of the IP-SF (5 items per area). Its selection criteria include item response theory discrimination, fidelity to the hexagonal structure, content coverage and, explicitly, "minimising gender differences within the RIASEC scales" (Rounds et al., 2016). The source report states that the Mini-IP is suitable for "being embedded in a longer multi-measure survey"; this is direct support for the two-stage design. In the data file each item carries the field `mini: true/false`; the content seeder writes this to the `is_core` column (`Pusula-API/database/seeders/ContentSeeder.php`), and `FormV4` asks only the `is_core` items in stage 1. The count gives 5 Mini-IP and 5 complementary items in each RIASEC area. That first-letter agreement between the two forms stays at κ = .73 is the source-level sign that the first letter of the pre-sign-up Holland code may change after sign-up; this observation is the background to the "neck-and-neck" decision in Chapter 4.

3.3.3 Work value cards

The work values pool is the 20 cards of the O*NET Work Importance Locator (WIL). The cards are statements that complete the stem "In my ideal job, it is important that…" ("…I make use of my abilities.", "…I would be treated fairly by the company.") and correspond to the needs in Dawis and Lofquist's Theory of Work Adjustment (`work_values.json`). The person distributes the cards over five columns ("Most important" 5 … "Least important" 1) so that there are exactly four cards in each col-

umn. The cards are tied to the six values in unequal numbers: Achievement (achievement) 2, Working conditions (working_conditions) 6, and Independence, Recognition (recognition), Relationships and Support three cards each (count from the file). This distribution is identical to the WIL worksheet (WIL instructions).

The reliability of the WIL is weak at source, and the reason for this is largely ipsative (within-person) scoring: in the computerised version the value-level retest over 4–8 weeks is r = .59–.66, and alpha ranges between .50 (Relationships/Altruism) and .86 (Independence/Autonomy); the top-ranked value remains first in the second administration in 62–70% of cases and remains in the top two in 88% on average (McCloy et al., 1999b). That O*NET carries only the two highest values into the score report is the design response to this weakness, and Rehnüma follows the same rule (3.7.4). The note in the data file states that the WIL has been officially retired and that the item texts were taken from the archived manual.

3.3.4 Rhythm questions

For rhythm (chronotype), the most widely used short scale in the literature, the rMEQ (five items, 4–25 points), was not used because its commercial licence is unclear (ADR-026). In its place Rehnüma's own three preference questions are asked: at what time the person would want to wake up if there were no obligations, at which hours they focus best on a difficult task, and at what time they would go to bed if there were no obligations. The five options of each question are scored between +2 (earliest) and −2 (latest) (`rhythm.json`). These three questions are not a psychometric scale, and the result is used in the report only in the language of suggestion (`RhythmScorer`). The literature supports this caution: the agreement of a single-item self-assessment with a behavioural indicator of chronotype is only "moderate" (weighted κ = .39; `turetilmis_katmanlar.md` §8), and cut-off points are population-sensitive.

3.3.5 Attention-check items

The three instructed-response items are presented on the same scale as the other items in the form and ask the person to mark a particular option: in stage 1, "For this item, select option 2, 'Not much like me'."; in stage 2, option 4 ("Quite like me") in the personality block and option 2 ("Dislike") in the interest block. All three checks deliberately ask for a non-extreme option, because in the data of Curran and Hauser (2019) an instructed-response item asking for "strongly agree" produced 8% false positives and one asking for "agree" 2%. The checks are written without the prefix "Attention check:"; announcing the check may increase purposeful non-compliance (Lundmark et al., 2023). On the other hand, the sentence "A few items only check that you are paying attention." has been added to the end of the general instruction: this is intended to remove the sense of a hidden trap and is different from announcing the instructed-response item itself (ADR-026, Addendum). There is evidence that instructed-response items do not harm scale validity or measurement invariance (Kung et al., 2018); three checks in 183 Likert items is consistent with the sparing use found in the literature.

So that they do not collide with the real items in the database, attention-check items receive negative identifiers and are marked with `code: 'X'`; the scorers skip this code. The test separately checks that the attention-check items do not enter the interest score (`test_her_asamada_dikkat_kontrolu_-var_ve_ilgi_puanina_karismaz`).

3.4 Translation and adaptation

3.4.1 Starting point and review method

Neither the IPIP-NEO-120 nor the O*NET Interest Profiler has a published Turkish adaptation. The only Turkish entry on the IPIP site is the 924-item pool translation by Somer, Korkmaz and Tatar; this pool contains only the Turkish sentence and the IPIP code, does not follow the IPIP-NEO structure and has not been verified by the IPIP project (Somer et al., 2002; IPIP translations page). The only point of comparison for RIASEC in Turkey is the 41-item adaptation by Süerdem and Erkök (2016); experts criticised the "word-for-word translation", the machinery items were perceived as men's work, the first confirmatory factor analysis gave RMSEA = .147, and the authors concluded that the scale needed to be revised substantially.

The Rehnüma translation was made by the developer and underwent a systematic item review on 22–23 October 2026 (`ceviri_ve_madde_kalitesi.md`). The review compared each item with its English source, with its counterpart in the Somer-TR pool (by matching on meaning; without code verification) and with Johnson's (2014) item–total correlations, and classified the problems into eight types: meaning shift, idiom/set-phrase translation, readability for age 14, double-barrelled item, added intensifier, overly extreme wording, cultural adaptation and repetition/near duplicate. ADR-029 §3 decided to apply the results of this review before the pilot.

3.4.2 Changes applied

As a result of the review, 34 IPIP items, 10 interest items and 2 WIL cards were rewritten (ADR-029 §3; note in `ipip_neo120.json`). The pattern of the changes can be grouped under four headings.

Meaning shift. The five most critical shifts carried facet content over into a neighbouring domain. The item "Take charge" had been translated as "Sorumluluğu üstlenirim" ("I take on the responsibility"); this measures C3 Dutifulness (sense of responsibility) content instead of E3 Assertiveness and was corrected with the proposal "İşlerin başına geçerim" ("I take the lead of things"). In the item "Wait for others to lead the way", the idiom "yol göstermek" (literally "to show the way") means giving advice in Turkish, so that the sense of leading was lost. The translation "işleri ağırdan almak" ("to take things slowly, to drag one's feet") of the item "Like to take it easy" evoked laziness (C−) instead of E4 Tempo. The translation "gözü kara" ("fearless, daring") of the item "Enjoy being reckless" carried the sense of heroic courage instead of negative recklessness. And "temas" ("contact") in the item "Avoid contacts with others" evoked physical touching in Turkish (`ceviri_ve_madde_kalitesi.md` §1).

Overly extreme wording. Expressions such as "intikam alırım" ("I take revenge") and "aşağılarım" ("I humiliate [people]") in the facets A2 Straightforwardness and A4 Compliance would have produced a floor effect and low variance in facets that are already entirely reverse-keyed; item A4:3 ("Insult people.") was brought into line with the Somer-TR rendering "hakaret ederim" ("I insult"). That the A5 Modesty facet hardly loads on the domain at all even at source (.05–.09) has been explained by social desirability (Kajonius & Johnson, 2019); that expressions of the type "kendimi çok beğenirim" ("I think very highly of myself"; original: "Think highly of myself.") may be culturally more off-putting in Turkish was noted as an inference.

Readability. Intensifiers ("hep", always; "çok", very; "bazen", sometimes) were removed; words of low frequency for age 14 such as "kayıtsız" (indifferent), "erdem" (virtue), "harfiyen" (to the letter), "kuramsal" (theoretical) and "ayartma" (temptation) were replaced; double negatives ("dışına çıkmamayı", "not stepping outside of") were unpacked. Before the review the 120-item set, at a mean of 3.18 syllables per word, was above Ateşman's threshold of 3.0 for the "most difficult text", and the Ateşman score at set level was 61 (Medium); the interest items scored 78 (Easy) and the WIL cards 66. The target for the revised set was put at Ateşman ≥ 70 and syllables per word ≤ 2.9; re-running the calculation on the revised text is open work. Formula scores are for relative comparison only; the data of Soto and colleagues (2008) showed that the real adolescent risk is not word comprehension but acquiescence bias (3.5).

Rewriting the O6 facet. The Dutch and Mexican Spanish translations of the IPIP-NEO-120 translated the voting and crime items of O6 literally; Rehnüma chose to rewrite instead. This choice rests on Johnson's (2014) characterisation of O6 as "the only scale whose content similarity is questionable", on the over-representation of crime and order items, and on the unsuitability of party/voting items for age 14 and for the Turkish context. The review had also found that the rewritten item O6:2 was a reversed copy of O6:4: asking the same content in two directions inflates alpha artificially and produces a correlation with the C3 item "Kuralları çiğnerim" ("Break rules."). O6:2 was therefore moved to the theme of the changeability of values, in the form "Toplumun değer ve kurallarının zamanla değişmesini doğal karşılarım." ("Find it natural that society's values and rules change over time."). The facet now has two positively keyed (O6:1, O6:2) and two reverse-keyed (O6:3, O6:4) items. If the alpha of O6 comes out below .60 in the pilot, hiding the facet in the report while maintaining its contribution to the domain score has been recorded as the fallback plan.

In the interest items the centre of the problem was not the Realistic area but the office and finance jargon in the Enterprising and Conventional areas: "tahvil" (bonds), "müvekkil" (a lawyer's client), "el terminali" (handheld terminal), "rehabilitasyon terapisi" (rehabilitation therapy) and a mail item with three verbs. Two cards were corrected in the WIL: the calque "maaşımın benzer işlerdekilerle iyi kıyaslanması" (a word-for-word rendering of "my pay would compare well with that of other workers") and a low-frequency verb such as "arkalamak" ("to back [someone] up"). Under §3(b) of the O*NET licence every change has to be written into a list of changes open to the user; this list and the English notice on the entry screen were completed with ADR-031 §5 (the licence obligations are treated in Chapter 8). The three rewritten O6 items have a second consequence: because these items are not present in the published open data of the scale, no percentile can be given for the O6 facet (3.10).

3.4.3 Response scale labels

The change of labels may be more important than the change of items. The old personality sequence "Hiç uymuyor / Pek uymuyor / Kararsızım / Uyuyor / Tamamen uyuyor" ("Does not fit at all / Does not fit much / I am undecided / Fits / Fits completely") was asymmetric (two graded steps on the negative side, an ungraded step on the positive side), and the midpoint labelled the person's state ("kararsızım", "I am undecided") rather than the applicability of the item. Doğan, Gündeğer and Yurtçu (2024) found only configural invariance with four different midpoint labels on the same 14 items and could not establish metric invariance; that is, a change of label changes the meaning of the scores. The new sequence is symmetric and person-focused; the context "Bana … uyuyor" ("… fits

me") also removes the homonymy of "Uyuyor" on its own with the verb "uyumak" (to sleep). The interest midpoint was made "Emin değilim" ("Unsure"), remaining faithful to O*NET's "unsure" option (ADR-029 §3).

The sequencing decision is binding from a measurement standpoint: the labels were changed *before* the pilot and will not change after it. Changing the labels after the pilot would have made the Gate 1–2 data incomparable with the data to be collected under the new labels.

3.4.4 Adaptation steps still open

The ITC guidelines for test adaptation (ITC, 2017) call for at least two independent forward translations and reconciliation, cognitive pretesting and a pilot item analysis with about 100 people; back-translation alone is not considered sufficient. The guidelines provide no special exception for the developer making the translation personally. According to the inference of the research note, the minimum acceptable route is a second independent translation, a record of reconciliation with a third person, and cognitive interviews with 5–10 people, at least three of them aged 14–17 (`dogrulama_protokolu.md` §1). These three steps have not yet been carried out; the present adaptation is a single-translator draft plus a systematic desk review. This state of affairs limits the language of the report in the way described in 4.8.

The test is offered today in five languages (Turkish, English, French, German, Arabic); the source language is Turkish and the content is read from a file per language (`FormV4::content`; ADR-046 §5). The English of the IPIP-NEO-120 and O*NET items is the original text; the French, German and Arabic item translations are new translations, have not been validated and have not undergone native-speaker review (ADR-046 §6). The translation review described in this chapter covers the Turkish items only.

3.5 Keying balance

3.5.1 The problem

Rehnüma's personality set is identical to Johnson's (2014) original in keying direction: 55 of the 120 items are positively keyed and 65 reverse-keyed, and nine facets are entirely unidirectional. All the items of C1, E5, E6 and O1 are positively keyed; all the items of A2, A4, A5, C6 and S1 (N1 in the NEO) are reverse-keyed (ADR-030; count from the data file). Johnson deliberately did not treat balanced keying as a criterion and wrote that he saw no convincing evidence that response biases play an important role in personality measurement. The domain-level distribution is given in Table 3.4.

Table 3.4. Domain-level distribution of keying direction in the personality items (positive / reverse).

Domain	Full set (120)	Stage 1 core (60)	Unidirectional facets	Number of bidirectional facets
S Emotional stability	7 / 17	5 / 7	S1 (all reverse)	5
E Extraversion	18 / 6	8 / 4	E5, E6 (all positive)	4
O Openness to experience	12 / 12	7 / 5	O1 (all positive)	5
A Agreeableness	7 / 17	3 / 9	A2, A4, A5 (all reverse)	3

Domain	Full set (120)	Stage 1 core (60)	Unidirectional facets	Number of bidirectional facets
C Conscientiousness	11 / 13	6 / 6	C1 (all positive), C6 (all reverse)	4
Total	**55 / 65**	**29 / 31**	**9 facets**	**21 facets**

Note. The numbers were computed from the `keyed` and `core` fields in the file `Pusula-API/database/data/ipip_neo120.json`. Because in the S domain the keys have been reversed from the N direction, the positive/reverse counts in Johnson's N table are swapped here; the absolute balance does not change.

The measurement consequence of the imbalance arises through acquiescence bias (the tendency to "agree" irrespective of content). In the large internet sample of Soto and colleagues (2008) aged 10–20, acquiescence varied widely among young people and strongly affected the psychometric properties of the scales; in 12,987 Brazilian children and adolescents fully balanced facets markedly improved criterion validity (Primi et al., 2020), and in 18 countries correction for acquiescence increased the congruence of factor loadings (Rammstedt et al., 2013). A concrete prediction follows from Table 3.4: a person who "agrees" one step more with every item scores high in the E domain (18 positively keyed items against 6 reverse) and low in the S and A domains (7 positive against 17 reverse). Part of a profile that appears in the report as "extraverted but anxious and disagreeable" may for this reason be a method artefact (ADR-030, Question).

3.5.2 Decision: measure, do not correct

ADR-030 chose among three options. Adding one reverse-keyed item from the IPIP pool to each of the nine unidirectional facets would have established balance but would have separated the item set from Johnson's form and forfeited direct comparison of the Gate 1–2 results with the published facet alphas and loadings. Applying within-person centring (ipsatisation) to every test would have been a new and arbitrary threshold in the absence of norm and distribution information. Removing the reverse-keyed items, in turn, would have lowered facet reliability and removed the basis of the consistency check. The decision is as follows: before the pilot the item set does not change (apart from O6), acquiescence is measured and stored in every test, and the score is not corrected.

The measurement is made by the function `Quality::acquiescence` (`Pusula-API/app/Scoring/V4/Quality.php`). The index is computed only from facets that have both positively and reverse-keyed items (21 facets in the full set and also 21 in stage 1): in each facet the mean of the raw mean of the positively keyed items and the raw mean of the reverse-keyed items is taken (*without reversing* the reverse-keyed items), the facets are combined with equal weight, and the scale midpoint (3) is subtracted. Because the content is represented equally in both directions, the person's true trait level is damped out in this mean; what remains is the tendency to agree irrespective of content. The index requires at least eight balanced facets; $|\text{value}| \geq 1.0$ produces the `acquiescent` flag.

This form had a preliminary version, which explains why the design is as it is. The first version computed the index over all items; this leaked the domain-level keying imbalance into the index, and in the simulation the index correlated $r = .33$ with E and $r = -.39$ with A. In the version restricted to balanced facets, in a re-run with 98 people, the correlations fell to $-.04$ for E, $-.06$ for O, $-.07$ for C, $-.24$ for A and $-.26$ for S; the trace remaining in A and S is the share of the domains with few bidi-

rectional facets and is below the .30 threshold (ADR-030 §2). In the simulation the index was at most 0.57 in honest synthetic persons and 2.0 in straightlining responders; the threshold of 1.0 separates these two groups (ADR-030, Rationale).

3.5.3 Decision point

ADR-030 §3 writes numerically the condition under which the decision will be reopened: if in the pilot ($N \approx 100$) the proportion with $|\text{index}| \geq 1.0$ in the 14–18 age group exceeds 10%, or if the correlation between the index and the E, S or A domain scores comes out at $|r| \geq .30$, the ADR is updated and either one reverse item is added to each of the nine unidirectional facets, raising the facet to five items, or within-person centring is carried out with the index computed from the balanced facets. The calculation that checks this condition is in the function `acquiescence_report` in `psychometrics.py` (4.7). The research report had recommended that the item change be made *before* the pilot; ADR-030 chose to measure, on grounds of comparability. This tension is discussed in Chapter 9.

3.6 Response-quality indicators and levels

3.6.1 Principle: warn without blocking

Careless responding is real and is frequent among adolescents: in undergraduate samples about 10–12% severe careless responding has been reported (Meade & Craig, 2012), and among 12,578 secondary-school students 14% of participants were flagged by at least one indicator under a lenient threshold and 37% under a strict one (Chauliac et al., 2023). Even so, the form does not block the result in case of inconsistency (ADR-026 §5). Active warning has increased drop-out, and blocking rewards *appearing* consistent rather than the true trait (research report; inference). The indicators are therefore reflected as an optional review screen and a confidence note. Only in a narrowly defined severe case is a "cannot be interpreted" result given instead of the narrative report.

3.6.2 Five indicators

`Quality::evaluate` computes five indicators at every completion and at the review check (Table 3.5). The indicators are an adaptation of the package recommended in the literature for joint use; no screening is done on a single indicator (Meade & Craig, 2012).

Table 3.5. Response-quality indicators, method of computation and flag thresholds.

Indicator	Flag	Computation	Threshold	Basis
Consecutive identical responses	`longstring`	The longest run of the same answer in order of presentation within a section; the maximum across sections	≥ 10	Rule originating with Huang et al. (2012), secondary citation; 8–10 in the practitioner guide
Instructed-response item	`check_failed`	An answer different from the expected one to any attention-check item	1 error	Curran & Hauser (2019); Kung et al. (2018)
Odd–even consistency	`inconsistent`	The key-corrected items of each facet are split into two halves (positions 0, 2 / 1, 3); Pearson correlation of the half means across facets; at least 10 facets	r < .30	Originating with Jackson (1976); as reported by Welz and Alfons (2023)

Indicator	Flag	Computation	Threshold	Basis
Page speed	`too_fast`	Proportion of pages that took less than 2 seconds per item	≥ 50%	Rule originating with Huang et al. (2012), secondary citation
Acquiescence	`acquiescent`	Mean of the positive and reverse raw means in balanced facets − 3; at least 8 facets	|value| ≥ 1.0	Soto et al. (2008); ADR-030

Note. Most of the thresholds come from secondary citations and have not been calibrated on Turkish or mobile data (`dogrulama_protokolu.md` §6, "Gaps"). Source: `Pusula-API/app/Scoring/V4/Quality.php`. ADR-026 §5 states "four indicators"; the fifth indicator (acquiescence) was added with ADR-029 §4.

Four of the five indicators are computed only from the answers of the Likert sections. Page speed is the exception: at completion the time records of all sections in scope enter the indicator, and in form v5 each round of statement ranking reports its time as a page of eight items (`AssessmentScorer::pages`; `SectionRunner.tsx`). A round ranked in less than 16 seconds is therefore counted as a "fast page"; the 2-second threshold was chosen for a Likert item, and its suitability for a ranking round has not been tested (inference from the code). Two further details of the code matter in the definitions of the indicators. First, longstring is counted in order of presentation; since the order within a section is shuffled and stored when the form is created, this is the order the person sees. The run is reset at a section boundary and at an unanswered item. Because reverse-keyed items are scattered in mixed order, ten consecutive identical answers are rare in a genuine profile (`dogrulama_protokolu.md` §6, inference). Second, the odd–even correlation may be undefined: if the same half mean results in all facets (a flat profile), the variance is zero and the function returns `null`; this case is not counted as inconsistency, because the flat profile is caught by longstring in any case. In stage 1 each facet has two items, and the two halves correspond exactly to the items in blocks A and B; this is the function of the A/B split in 3.2.

3.6.3 Levels

The indicators are converted into a four-level decision. The code determines the level in the following order: if an instructed-response error is seen together with longstring, or if 99% or more of the pages are excessively fast, the case is *severe* and the level becomes `uninterpretable`; otherwise two or more flags receive the level `review`, a single flag `note`, and a result without flags `ok`. The `reliable` field is false only at the `uninterpretable` level (Table 3.6).

Table 3.6. Response-quality levels, their conditions and their effect on the user flow.

Level	Condition	Review screen	Report	`reliable`
`ok`	No flag	None	Normal	true
`note`	Single flag	None	Normal; the flag is kept in the technical detail	true
`review`	Two or more flags	Opens optionally if there is a contradicting pair	Given, with a confidence note	true
`uninterpretable`	Instruction error + longstring ≥ 10, or fast-page proportion ≥ 99%	Opens optionally if there is a contradicting pair	"Cannot be interpreted" screen and an invitation to retake, instead of the narrative report	false

Note. There are four levels in the code: `ok`, `note`, `review` and `uninterpretable` (`Quality::LEVELS`, `Quality::evaluate`); there is no level named `caution`. With form v5 the level is converted into a readable label in the report (ADR-044 §4; `Quality::labels`): the Turkish texts are, in order, "Cevapların tutarlı" ("Your answers are consistent"), "Bir nota bak" ("One thing to note"), "Gözden geçir" ("Worth reviewing") and "Yorumlanamaz" ("Cannot be interpreted"), and each comes together with its one-sentence reason (`Pusula-API/lang/tr/messages.php` → `quality`). The levels at which the review screen opens are determined in the front end (`Front-end/src/components/test/TestFlow.tsx`: level `review` or `uninterpretable` and at least one contradicting pair).

The code agrees with the definition in ADR-026: "two or more indicators: optional review screen; cannot be interpreted only in the severe case (excessive speed on every page, or instruction + longstring)". The code operationalises the expression "on every page" as a proportion of ≥ 99%.

3.6.4 The review screen and the single-item rule

The content of the review screen comes from the function `Quality::reviewPairs`: item pairs in the same facet that differ by at least 3 points after key correction are sorted by the difference, and at most six of them are returned. A typical example is answering "Exactly like me" to both of the items "Get angry easily." and "Am not easily annoyed." (Turkish: "Kolayca sinirlenirim", "Kolay kolay sinirlenmem"). `AssessmentScorer::check` adds the item texts and the previous answers to these pairs; because the identifiers of personality and interest items come from separate tables, the text is found by using section and identifier together.

The first design of the screen showed the two items of the pair side by side. The addendum to ADR-026 dated 23 October 2026 changed this: showing the two items side by side exposed the reverse-keying structure and created a demand for consistency in the person. In that case the person corrects not their trait but their contradiction; internal consistency is inflated artificially and the Gate 1–2 data are contaminated. The screen now asks again **a single item from each pair**, in shuffled order and without showing the previous answer; items that are not answered remain as they are. In the front end the selection has been implemented as taking the second item of the pair, de-duplicating where the same item occurs in two pairs, and ordering by a fixed identifier-based shuffle (`Front-end/src/components/test/steps/ReviewPage.tsx`). The text of the screen says explicitly that there is no need to change anything ("Sometimes that is simply how it is; you don't need to change anything."), and the option of proceeding without changing ("Carry on as I am") is offered with equal weight.

The rule is a trade-off: the answer given again is not an observation independent of the first answer; the single-item rule only prevents the correction from being triggered by a visible contradiction (inference). The proportion and direction of the answers that change after review should be reported separately at Gate 1 (4.7).

3.7 Scoring

Scoring is carried out entirely on the server and on the raw answers; no score or result is accepted from the client (`AssessmentScorer` class comment). The common transformation in all Likert scales is the function `Likert::toPercent`: the 1–5 mean is converted to the range 0–100 by (mean − 1) / 4 × 100 and rounded to one decimal place (`Pusula-API/app/Scoring/Likert.php`). This scale is not a percentile; it is a linear re-expression within the scale's own range. A score of 50 means that the items

were answered on average with the middle option (3); it carries no normative sense of an "average". This distinction is the reason why in the report the percentile is given from a separate source, the published norm table of the scale (3.10).

3.7.1 Big Five domains and facets

`PersonalityScorer` computes the score of the five domains: in reverse-keyed items the answer is taken as 6 − answer, and the mean of all answered items of the domain is converted to 0–100. A domain with no items at all receives 50. The domains are ranked with `Ranking::descending`; for equal scores a fixed order (O, C, E, A, S) is preserved, so that a tie does not systematically favour any type and the same answers always give the same result (`Pusula-API/app/Scoring/Ranking.php`). `FacetScorer` performs the same operation at facet level and returns for each facet the score, the number of items the score comes from (`items`: 2 in stage 1, 4 in stage 2) and a three-band label: ≥ 60 high, ≤ 40 low, in between medium. The item-count field determines the confidence language of the report (ADR-026 §6): a two-item facet is marked as a "first impression".

Reporting the facets in only three bands is the product counterpart of Johnson's warning about the inadequacy of the facet scales for individual decisions. The most common convention in the literature is ±0.5 SD relative to the norm group (approximately bottom 30% / middle 40% / top 30%); however, in Johnson's (2020) calibration study with 160 people this cut gave only 41% correct classification against acquaintance ratings, and the optimal cuts were found about 1.2–1.3 SD from the mean. Because the Turkish version has no norm, Rehnüma's 40/60 band is defined on the raw scale, and there is no published source for a band of the form "raw score ± fixed difference" (`turetilmis_katmanlar.md` §4); the bands therefore remain in the language of "tendency".

3.7.2 RIASEC interest scores and expressed interest

`InterestScorer` converts the item mean of each RIASEC area to 0–100. The scores are normative (independent of one another): the Investigative scores of two people can be compared directly. The Holland code is the letters of the three highest areas; in a tie the fixed order R-I-A-S-E-C is preserved. This agrees with O*NET's own practice: the three letters are the three highest totals, and on the paper form in a tie the person is told to "choose the one that fits you best" (O*NET IP manual; `turetilmis_katmanlar.md` §6).

Expressed interest (the "most 2 + least 1" selection) does not enter the score. `ExpressedInterest::apply` does only two things. The first is tie-breaking: if in the measured ranking the score difference between two adjacent areas is 5 points or less (`TIE = 5.0`) and the lower area is higher in the person's expressed preference (most-selected +1, least-selected −1, others 0), the two areas swap places; the scan is repeated for as many passes as there are areas. The Holland code is regenerated from this new order; the measured ranking is stored as `measured_ranking`. The 5-point threshold was chosen as the counterpart on this scale of the "rule of 8" of the Self-Directed Search; in the data of Glavin and Savickas (2011), a code difference above measurement error raised the hit rate for choice of major from 40% to 55% (research report). It has been noted that the "rule of 8" is not found in official O*NET documents and comes from the SDS (`turetilmis_katmanlar.md` §6). The second is the agreement reading: if both of the two areas the person selected most are among the measured top two areas and the area selected least is among the measured bottom two areas, the label

`aligned` is given; if at least one of the most-selected is in the measured top two, `partial`; otherwise `different`. This is the counterpart of the SDS logic of interpreting the aspiration code alongside the measured code.

The same function produces the `top_two_close` flag: if the difference between the top two areas after tie-breaking is 5 points or less, the first letter of the Holland code is counted as "neck-and-neck". This flag arose directly from the Gate 0 finding (4.4–4.5).

3.7.3 The stage logic of scoring

When stage 1 is completed, the five domains are scored from 60 items, the facets from 2 items each and the RIASEC areas from 5 items each; no values or rhythm score is produced, and of the ranking steps only self-placement is present (statement ranking is in stage 2). In stage 2 all sections are scored again and together; the facets are computed from 4 items each and the RIASEC areas from 10 items each. The result is written to the same record (`assessment_results`) on both occasions and carries the scoring version in the `scoring_version` field. The `provisional` field in the profile object is true in stage 1 and is tied to the "first impression" language of the report (`ProfileBuilder::build`).

3.7.4 Work values and rhythm

`ValuesScorer` takes the score of each card as the number of the column in which it was placed (5 … 1), converts the mean of the cards of each value to 0–100 by (mean − 1) / 4 × 100, and ranks the six values; the first two values are brought forward as `top`. The scoring is ipsative: only the within-person ranking is meaningful, and the value scores of two people are not compared. The WIL instructions use multipliers to bring each value to the equivalent of six cards (Achievement ×3, Working conditions without a multiplier, the others ×2; value range 6–30). In the code the card mean is used instead of the multiplier. Because total × (6 / number of cards) = 6 × mean, the two methods are linear transformations of each other and give the same ranking (inference; in terms of outcome this does not conflict with the research note's recommendation that "the multipliers should be used directly").

`RhythmScorer` sums the option scores of the three questions (−6 … +6); a total ≥ 3 is labelled morning type, ≤ −3 evening type, and in between intermediate type.

3.7.5 Self-placement and statement ranking

The two steps of form v5 are scored separately and enter no Likert score (`Pusula-API/app/Scoring/V5/`). `SelfPlacement::apply` only records: it returns the keys of the three archetypes the person selected, in order of selection (`closest`), and the archetype they found farthest (`farthest`); it produces no number.

`RankScorer::score` assigns points to the rank in each round: the statement in first place receives 8 points and the one in eighth place 1. The points are summed per archetype over the eight rounds (8–64 points per archetype; the total is always 288), and the archetype's share is written as a percentage of the total; as in archetype similarity, the percentages are completed to exactly 100 by the largest remainder method. The result is a within-person profile: the percentages are dependent on one another and produce no between-person comparison or percentile. The function also returns an informational consistency value (`consistency`, 0–1): the proportion of rounds in which the statements of the archetypes ranked in the top three overall were ranked in the upper half (5 points and above). No threshold

or flag is defined for this value, and it does not enter the quality level. The test checks, on a two-round example, that the points are assigned according to rank and that the percentages sum to 100 (`FormV5Test`: `test_siralama_puani_yuzde_yuz_toplar_ve_siraya_gore_puanlar`).

Note. ADR-044 §2 describes the score as "the total per archetype, divided by the number of rounds and converted to a percentage"; in the code there is no division by the number of rounds, and the share is taken directly as a proportion of the grand total. The two descriptions give the same ranking; here the code has been taken as authoritative.

3.8 Derived layers

The readable layers of the report make no new measurement; they convert the measured Big Five domain and facet scores (and, in one layer, the RIASEC scores as well) into labels and texts by rules. This work is done by `Pusula-API/app/Profile/ProfileBuilder.php`; the texts are in the file `database/data/report_texts.json`, and the class only selects and orders. The principle of ADR-028 is applied to all of these layers: "the label is the door, the content is the person". The visible language is motif, temperament and archetype; but the person-specific lines of each section come from the facets, so that two people who receive the same label do not read the same text. The one layer that stands apart is the three-way comparison: it places the derived archetype order side by side with the two ranking steps of form v5 (3.8.6). The basis in the literature and the limits of each layer differ; Table 3.7 summarises from the code the derivation rule and the thresholds of these layers.

Table 3.7. Derived layers: input, rule and thresholds.

Layer	Input	Rule	Threshold and constants	Source code
Mind motif (4 letters)	E, O, A, C domains	E→E/I, O→N/S, A→F/T, C→J/P; first letter if score ≥ 50	Clarity: |score − 50| < 8 balanced, < 20 moderate, ≥ 20 clear	`ProfileBuilder::code`
Alternative motifs	Balanced letters	Each balanced letter is flipped one at a time; if there are two or more balanced letters the first two are flipped together	At most 3 alternatives	`ProfileBuilder::alternatives`
Temperament	Motif letters	If the 2nd letter is N, N + 3rd letter (NF, NT); otherwise S + 4th letter (SJ, SP)	—	`ProfileBuilder::temperament`
Archetype (8)	Five domains	z = (score − 50) / 25; Gaussian weight by distance to the prototype; conversion to a percentage	σ = 0.8; percentages completed to 100 by the largest remainder method	`ProfileBuilder::archetypes`
Domain band	Five domains	≥ 60 high, ≤ 40 low	—	`ProfileBuilder::domains`
Facet band	30 facets	≥ 60 high, ≤ 40 low	—	`FacetScorer::band`
Strength / growth	30 facets	"Strength" or "cost" text according to the pole of facets with |score − 50| ≥ 15	At most 9; `certain` = facet measured with 4 items	`ProfileBuilder::facetSides`
Personal lines	Facets tied to the section	The two facets of the section that deviate most from 50	Deviation ≥ 15	`ProfileBuilder::personalLines`
Natural tendencies (10)	Facet and RIASEC scores	Weighted mean deviation: 50 + 50 × Σ w (s − 50)/50 / Σ w	Band ≥ 60 high, ≤ 40 low	`ProfileBuilder::tendencies`

Layer	Input	Rule	Threshold and constants	Source code
Three-way comparison	Measured archetype order, self-placement, statement ranking	Number of shared archetypes in the top threes of the sources, taken two at a time	0 divergent, 1 partly, 2 largely, 3 fully overlapping	`ProfileBuilder::triangulate`

Note. All thresholds are code constants (`BALANCED = 8.0`, `DISTINCT = 15.0`, `SIGMA = 0.8`) and, in the words of ADR-029, "not yet based on data"; it is planned to re-establish them with the Gate 2 and Gate 4 data.

3.8.1 Mind motif, the balanced threshold and alternative motifs

The basis of the four-letter code is the finding of McCrae and Costa (1989) showing that the four indices of the MBTI measure four dimensions of the Big Five: E–I with Extraversion $r = -.74$, S–N with Openness $r = .72$, T–F with Agreeableness $r = .44$, J–P with Conscientiousness $r = -.49$ (the signs depend on the direction of the MBTI score). The same study found no support for "truly dichotomous preferences"; MBTI scores are distributed not bimodally but with a peak at the centre, and a change of type on retest after five weeks has been reported in the range of 39–76% (as reported by Pittenger, 1993). The T–F and J–P letters are for this reason structurally noisier. The code derives the letters solely from the position of the domain score relative to 50, the scale midpoint; Neuroticism (emotional stability) does not enter the code. The report text carries everywhere the note "This code is derived from Big Five scores; it is not related to Myers-Briggs (MBTI).", and the names and texts of the 16 motifs are Rehnüma's original content (ADR-028 §6).

The cut-off point is not the norm median but the scale midpoint; this is a provisional solution for the period in which there is no norm (research report). The clarity of each letter is read from the distance of the score from 50. The balanced threshold was initially 6 points; when the Gate 0 simulation showed that in the retest clear letters never flipped whereas about one third of balanced letters did, the threshold was moved to 8 points (ADR-029 §2). The rationale is that the standard deviation of the retest difference is about 6–7 points; the 8-point band covers the letters for which the side of the boundary on which they fall in a single administration lies within measurement error. The closest basis in the literature is that, where reliability is assumed to be .80, a score with $|z| \leq .78$ cannot be said at the 5% level to differ from the mean (Gerlach et al., 2019).

A balanced letter is not hidden from the report: the `alternatives` field returns the neighbouring motifs (at most three) obtained by flipping these letters; if one letter is balanced one alternative results, and if two letters are balanced three (the single-letter case is tested by `test_dengede_harf_komsu_motifleri_alternatif_olarak_verir` in `ProfileBuilderTest`). In this way the flipping of a single letter at the boundary does not change the report entirely; the fragility of the categorical label is not hidden but carried along. The ITC's requirement that classification accuracy be documented in categorical reports ties this design to the stability measurement in Chapter 4.

The motif has four sections in the report (Overall pattern, Your motif at work, Typical pursuits, Your motif and relationships); to the base text of each section are added at most two personal lines coming from the facets of that section. The section–facet mapping has been transferred to the code exactly from ADR-028 §2 (`SECTION_FACETS`): work C1, C2, C4, C5, C6, E3, O5; pursuits O1, O2, O4, E4, E5; relationships A1, A3, A4, A5, A6, E1, E2, S4; under pressure S1, S2, S3, S5, S6. The personal lines of

the "Overall pattern" section come not from the facets but from the two most pronounced of the five domains that fall outside the 60/40 band. If no facet has moved 15 points away from 50, no line is written; the test verifies that no personal line is produced when there is no distinct facet.

3.8.2 Temperament

Temperament is the Big Five counterpart of Keirsey's four temperaments (Guardian SJ, Artisan SP, Idealist NF, Rational NT) and is derived from the motif letters: if the second letter is N, the temperament consists of N and the third letter (F or T); otherwise of S and the fourth letter (J or P). The partition is theoretical (Keirsey, n.d.). No evidence has been found that these four groups form empirical clusters in Big Five space, and because the two axes rest on different factors the partition is asymmetric: SJ ≈ low O + high C, SP ≈ low O + low C, NF ≈ high O + high A, NT ≈ high O + low A (inference). Temperament is therefore the layer with the weakest evidence grade and remains within the product as a narrative frame. It has six sections in the report (Overall pattern, What moves you, At work, In relationships, Under pressure, Direction of growth); the work, relationships and pressure sections take personal lines from the same facet mapping as the motif, and the growth section shows the "try" suggestions of the two most pronounced facets. The Turkish names are Rehnüma's own ("Anlam arayan", Seeks meaning; "Çözüm arayan", Seeks solutions; "Düzen kuran", Builds order; "Deneyerek öğrenen", Learns by doing); under ADR-032 §4, Keirsey's universal code and name are also written alongside in the report.

3.8.3 Eight archetypes

The archetypes rest on the finding of Gerlach and colleagues (2018): of the 13 clusters found with a Gaussian mixture model in four data sets of more than 1.5 million people, only four passed the "meaningful cluster" criterion (Average, Reserved, Role model, Self-centred). The basis is real but limited. The authors' 2019 reply states that assigning the individual to a type was not the aim, that they agree with the observation that only 42% of participants fell into a single type, and that they support the call to avoid typology in practice (Gerlach et al., 2019); Katahira and colleagues (2020) showed that a skewed unimodal distribution without cluster structure can produce "meaningful" clusters under the same null model. Moreover, the numerical z values of the cluster centres are not in the body of the article; the research note was unable to verify the supplementary table (`turetilmis_katmanlar.md` §1). Rehnüma's division of the four cores into eight is a product decision that goes beyond the source (ADR-026 §3), and the prototypes were written by hand. Table 3.8 gives the eight prototypes in z units.

Table 3.8. Prototypes of the eight archetypes (z units; S in the direction of emotional stability).

Key	Name	S	E	O	A	C	Relation to Gerlach core (inference)
`rehber`	Steady Guide	0.8	0.8	0.6	0.8	0.8	Role model
`gozlemci`	Quiet Observer	0.4	−0.8	−0.5	0.2	0.2	Reserved
`denge`	Flexible Balance	0.0	0.0	0.0	0.0	0.0	Average
`oncu`	Bold Pioneer	0.0	0.8	−0.3	−0.8	−0.6	Self-centred
`kasif`	Curious Explorer	0.0	0.3	1.0	0.0	−0.4	— (Rehnüma additions)
`mimar`	Meticulous Architect	0.3	−0.2	−0.3	0.2	1.0	—
`destekci`	Caring Supporter	−0.4	0.2	0.3	1.0	0.1	—

Key	Name	S	E	O	A	C	Relation to Gerlach core (inference)
`dusunur`	Deep Thinker	−0.1	−0.7	0.8	0.0	0.2	—

Note. The values were taken from the file `Pusula-API/database/data/report_texts.json` (`archetypes[].prototype`). The last column is the authors' interpretation of the mapping; the archetype–Gerlach cluster mapping is not explicitly tabulated in the data file or in the ADRs. ADR-029 has placed on record that the Gerlach centres could not be verified numerically and that our own centres will be derived by clustering the Gate 2 data.

Assignment is made on a soft-clustering logic. The person's domain scores are scaled by $z = (\text{score} - 50) / 25$; for each prototype the squared Euclidean distance d^2 is computed, and the weight is found by $\exp(-d^2 / (2\sigma^2))$ with $\sigma = 0.8$. The weights are normalised and converted to percentages; because independent rounding can make the total 99 or 101, the percentages are completed to exactly 100 by the largest remainder method (ADR-029, Impact). The report shows all eight archetypes with their similarity percentages; the first three are open, the others expand on demand (ADR-028 §3). This presentation is the closest to Gerlach's "soft clustering" logic and does not put the person into a single box.

The scaling calls for a caveat: the z here is not a score standardised against a norm group; it is a transformation in which 25 points on the raw 0–100 scale count as one unit. A person's closeness to the "Steady Guide" rests on the assumption that the prototype's values of 0.8 z correspond to 70 points on the raw scale (inference). When a norm has been formed, the prototype distance will need to be recomputed with true z scores.

3.8.4 Strengths and growth areas

Strengths and growth areas are derived from the rules of the 30 facets; their number is not fixed and varies with the profile (ADR-026 §3). The function `facetSides` finds the facets that have moved at least 15 points away from 50, selects the "high" or "low" pole of the facet according to the direction of the score, and returns the "strength" or "cost" text of that pole (fields strength and cost) and, on the growth side, additionally the "try" suggestion (tip); the list is sorted by distance and contains at most nine elements. Each element, as an expandable card, carries the sections "how it looks in daily life", "why in you" (score) and "try" (ADR-028 §3). The `certain` field says whether the facet has been measured with four items; the two-item facets in stage 1 are marked as a "first impression" by means of this field.

Two design features of the rule are a safeguard against the Barnum effect. First, both poles are written as strengths: a low Orderliness (C2) score has a strength too, and a high Orderliness score has a cost too; the note in the report texts file explicitly records the rule that "no sentence was written that would also be true for a person with the opposite score" (`report_texts.json`). Second, the list may remain empty if there is no distinct facet; there is no fixed 9/9 template. The research report recommended that, once a norm has been formed, the ±15-point threshold be combined with $|z| \geq .8$ and with Johnson's rule "if the subdomain score is in a different band from the domain score, attend to the subdomain". The facet band (40/60, that is, ±10) and the strength threshold (±15) are different: a facet may be in the "high" band and yet not enter the strengths list. This distinction is deliberate; the band is descriptive, the list selective (inference).

3.8.5 Ten natural tendencies

The competency layer is a derived layer that requires no items and is named "natural tendency", not "measured competency" (ADR-026 §3). The framework with the best evidence is Bartram's (2005) "Great Eight" model: in a meta-analysis of 29 studies, personality composites predicted all eight competencies at a moderate-to-good level, and personality and ability together gave an operational validity of .20–.44. The Big Five mapping of the O*NET work styles also appears in an official table (Putka et al., 2024). On the other hand, no direct evidence has been found that facet-level linear composites predict work-style ratings (`turetilmis_katmanlar.md` §5). The rationale for choosing ten tendencies instead of the 45 competencies of the competing product is that deriving 45 scores from five domain scores does not go beyond relabelling the same variance (ADR-026, Rejected alternative).

The score of each tendency is the weighted mean of the deviations of its components from 50: $50 + 50 \times \Sigma\, w_i\, (s_i - 50) \,/\, 50 \,/\, \Sigma\, w_i$. A component may be a facet or an interest area with the prefix `riasec:`; an unmeasured component drops out of both the numerator and the denominator. Table 3.9 gives the components and weights of the ten tendencies.

Table 3.9. Components and weights of the ten natural tendencies.

Key	Name	Components (weight)
`leadership`	Leadership and direction	E3 (1), C1 (0.5), S6 (0.5), RIASEC E (0.5)
`collaboration`	Collaboration	A4 (1), A1 (0.5), A3 (0.5), E1 (0.5)
`communication`	Communication and persuasion	E1, E2, E3, S4, RIASEC E (0.5 each)
`analysis`	Analytical thinking	O5 (1), C6 (0.5), RIASEC I (1)
`creativity`	Creativity	O1 (1), O2 (0.5), O4 (0.5), RIASEC A (1)
`planning`	Planning and order	C2 (1), C5 (1), C6 (0.5), RIASEC C (0.5)
`drive`	Results focus	C4 (1), C1 (0.5), E4 (0.5)
`resilience`	Calm under pressure	S6 (1), S1 (0.5), S2 (0.5), S3 (0.5)
`care`	Empathy and support	A6 (1), A3 (1), O3 (0.5), RIASEC S (0.5)
`integrity`	Reliability	C3 (1), A2 (1)

Note. Source: `Pusula-API/database/data/report_texts.json` (`tendencies[].components`). The weights were chosen by hand; they are consistent in direction with the mappings of Bartram (2005) and of Putka and colleagues (2024), but the coefficients do not come from a regression.

Because the tendency score is the mean of its components, it shrinks towards 50 when the components deviate in different directions; it is harder for the five-component Communication and persuasion tendency to take an extreme value than for the two-component Reliability (inference). The tendencies are shown with the 60/40 band and sorted by score; the report presents them in the language of "natural tendency", with the note that their predictive power is low to moderate.

3.8.6 Three-way comparison

With form v5 a comparison layer was added to the profile (ADR-044 §3). `ProfileBuilder::triangulate` places three sources side by side: the self-view (the three closest archetypes in the self-placement step), the expression profile (the three archetypes with the highest shares in statement ranking) and the measurement (the top three in the archetype similarity derived from

the Five-Factor domain scores; 3.8.3). For each pair of sources the number of shared archetypes in the top threes is computed (`self_measured`, `expression_measured`, `self_expression`; 0–3), and the number is converted into a label: 0 "divergent", 1 "partly overlapping", 2 "largely overlapping", 3 "fully overlapping". Order is not taken into account in the intersection; only presence in the top three is counted. Because statement ranking has not yet been done in stage 1, only the self-view–measurement pair exists; the other two pairs are added in stage 2. The function also marks whether the archetype the person called "farthest" enters the top three in the measurement (`farthest_in_measured_top`).

The comparison is not a decision rule: no intersection value hides the result, restarts the test or changes the quality level (ADR-044 §3); nor do the self-view and the expression profile enter the rule engine as conditions (ADR-044 §6). The text of the "divergent" label says this explicitly: the difference is presented not as an error but as a difference between the way the person sees themselves and what the measurement says (`Pusula-API/lang/tr/messages.php` → `profile.overlap`). The three sources are not independent: the archetype cards and the statements were written from the same eight prototypes, and there is as yet no distribution on which to base an interpretation of the level of overlap (inference); the thresholds of the labels are, as in the other derived layers, code constants not based on data. The test checks the intersection counts and the two extreme labels ("fully overlapping", "divergent") (`FormV5Test`: `test_uclu_karsilastirma_etiketleri`).

3.9 Occupation matching

Occupation matching compares the person's interest and personality profile with the profiles of the active occupations loaded from the O*NET 31.0 database (`Pusula-API/app/Matching/OccupationMatcher.php`). The occupation catalogue has 923 records (`Pusula-API/database/data/occupations.json`); 13 of these are marked `active: false` because they have no counterpart in Turkey and do not take part in matching, so the person is matched with 910 active occupations (count from the file; ADR-074 also says 910). The "923" in the configuration file and in ADR-046 is the number of catalogue records. All of the 910 active occupations have an interest profile and 878 have work-style data; the 32 occupations without work-style data are ranked by interest fit alone (3.9.2). The basic decision of the design is that interest is the main component and personality a relative adjustment, because interest has stronger evidence in occupational choice (class comment). Table 3.10 brings together all the parameters of the matching.

Table 3.10. Occupation matching parameters and their sources.

Parameter	Value	Location
Interest fit: weight of profile correlation	0.55 (negative r counted as zero)	`OccupationMatcher::interestFit`
Interest fit: weight of the Iachan M index	0.30 (weight matrix [[22,10,4], [10,5,2],[4,2,1]], maximum 28)	`OccupationMatcher::iachan`
Interest fit: weight of centrality	0.15 ((score of the person's top area in the occupation − 1) / 6)	`OccupationMatcher::interestFit`
Personality shift coefficient	6.0 (can be changed through the `model_parameters` table)	`config/pusula.php` → `parameters.personality_shift`
Personality shift limit	±8 points	`OccupationMatcher::MAX_PERSONALITY_SHIFT`

Parameter	Value	Location
Fit bands	very high ≥ 85, high ≥ 75, moderate ≥ 60, below that low	`config/pusula.php` → `fit_bands`
Confidence band: interest not distinct	interest spread < 25 → at most "moderate"	`report.min_interest_spread`
Confidence band: moderate distinctness	interest spread < 40 → at most "high"	`report.clear_interest_spread`
Number of stored suggestions	100	`report.stored_match_count`
Upper limit per family	8 (SOC major group)	`report.top_per_group`
Candidate pool factor	×3 (first 300 ranks)	`report.top_pool_factor`
"Challenging" occupations	5 (from the end of the fit order, those not in the suggestion list)	`report.stored_challenge_count`

Note. In the code ±8 is the *limit* of the personality shift; the coefficient that determines the size of the shift is 6.0. The two values must be read together.

3.9.1 Interest fit

Interest fit is a three-part composite. The shape component (55%) is the Pearson correlation between the person's six RIASEC scores and the standardised values of the occupation's O*NET interest profile; a negative correlation is counted as zero so that an "unrelated" occupation does not receive half the points. This is the essence of the O*NET Interest Profiler's own matching algorithm (correlation between the person profile and the occupation profile; "Best Fit" r ≥ .729) and automatically excludes profile elevation (O*NET IP manual). With correlation alone, hundreds of occupations were receiving a fit above 95%; the other two components were added to break this ceiling effect (code comment).

The code component (30%) is the Iachan (1984) M index between the three-letter Holland codes of the person and of the occupation. The index is sensitive to the order of the letters: if the person's first letter matches the occupation's first letter it receives 22 points, and if it matches the second letter 10 points, and the total is divided by 28 to bring it into the range 0–1 (implementation of the holland R package; `turetilmis_katmanlar.md` §6). The test verifies that RCI and RIC are not counted as the same (`test_iachan_index_rewards_letter_order`). In the literature the C index is considered the best method for this purpose (Eggerth & Andrew, 2006); that complex congruence indices do not predict satisfaction better (Hoeglund & Hansen, 1999) suggests that the effect of the choice of Iachan on the outcome will be small (inference). The centrality component (15%) is how decisive the person's strongest interest area is in that occupation, and it separates occupations that share the same code from one another.

If the interest profile is completely flat (all areas with the same score), standardisation is undefined and interest fit is not computed; in this case the ranking rests on personality fit, and the result is marked "not distinct" in any case. This is the counterpart of O*NET's invalid-profile rule of "strongly dislike to all items".

3.9.2 Personality shift

Personality fit is computed on the *shape* of the two profiles. First the occupation's "demand" for each Big Five domain is found as the mean of the O*NET work styles tied to that domain; the mapping is in the list of 21 work styles in `config/pusula.php` (e.g. Attention to Detail → C, Stress Tolerance → S, Cooperation → A). The demand is standardised across all occupations (every occupation asks for dependability; what matters is how much more it is asked for in this occupation than in the others) and centred within the occupation. The person's domain scores are likewise centred on their own mean; in this way someone who says "agree" to every statement does not appear suited to every occupation. The personality score is the covariance of the two centred profiles and is converted to 0–100 by $50 + 50 \times \tanh(\text{score})$.

The shift is proportional to the deviation of the personality score from the person's *mean across all occupations*: shift = 6.0 × (personality score in the occupation − mean), limited to ±8. Total fit = interest fit + shift, clipped to the range 0–100. The consequence of this form is that an occupation better suited to the personality rises, a less suited one falls, and an occupation without work-style data neither gains nor loses. The tests separately check these three behaviours and that a profile agreeing with everything does not change the ranking (`OccupationMatcherTest`).

3.9.3 Fit bands and the confidence band

Instead of a percentage the user is shown four bands: very high (≥ 85), high (≥ 75), moderate (≥ 60) and low. The configuration file states that the thresholds were chosen "according to the distribution over 923 occupations with sample profiles" (923 is the number of catalogue records; 910 active occupations enter the matching); that is, the thresholds rest not on an empirical criterion but on the distribution being discriminating. O*NET's own bands, by contrast, are tied to the significance level of the correlation ($r \geq .729$, one-tailed $p < .05$). Because Rehnüma's composite score is not a correlation alone, these thresholds cannot be carried over directly.

`FitConfidence` adds a confidence cap to the bands. The distinctness of the interest profile is measured as the difference (0–100) between the strongest and the weakest area in the Likert scores; if the difference is below 25, the profile is counted as "not distinct" and no occupation is shown higher than "moderate"; if it is below 40, the band can be at most "high". The ranking does not change; only the label is capped. The configuration note records that random answers generally produce a difference of between 15 and 30. Expressed interest does not measure distinctness; it only breaks ties (ADR-026). This rule is the in-product counterpart of the concept of "differentiation" in the literature, that is, the difference between the highest and the lowest RIASEC score (`dogrulama_protokolu.md` §3).

3.9.4 Selecting the stored list: `selectTop`

The whole of a person's matching with the 910 active occupations can be recomputed on every request (`AssessmentFits`); only the suggestion list is stored in the result. The fit of the target occupation optionally selected at the start of the test is likewise read from the same matching, not from a separate formula (ADR-074). The selection rule of the list was changed by the addendum to ADR-032. The old rule was "at most 3 occupations from the same family in the first 15 ranks, then the rest in order of fit"; this caused rank 16 to come out with a higher fit than rank 15 and was read as an error in a list shown with percentages. The new rule is in the function `OccupationMatcher::selectTop`:

the first 300 candidates ordered by fit (the 100 to be stored × pool factor 3) are traversed; at most 8 occupations are taken from the same SOC major group (the first two digits of the O*NET code; e.g. 47 construction, 29 healthcare); the traversal stops when 100 occupations have been reached; the selected ones are sorted again by fit. The result is monotonic (the fit of rank n is not lower than that of rank n+1) and diverse (in a Realistic profile the top ranks are not filled by a single family). Limiting the candidate pool to 300 prevents occupations whose family is under-represented but whose fit is low from leaking into the list; when the candidates run out, the list may remain shorter than 100. In the example profile of ADR-032 the list is 81 occupations long and its tail is at 74% fit. Occupations exceeding the capacity drop out of the list, but the full ranking can always be seen in the occupation catalogue. The tests check the family limit, monotonicity and the pool limit (`OccupationMatcherTest`).

The expectation in the literature regarding the validity of occupation suggestions is modest: the correlation between the suggested occupation and user liking is expected to come out small ($r \approx .10–.20$), and complex congruence indices do not increase it (Hoeglund & Hansen, 1999; `dogrulama_protokolu.md` §8). For this reason the parameters of the matching (weights, shift coefficient, band thresholds) are each a product decision and should be re-evaluated with the criterion data at Gate 3.

3.10 Percentiles against the published norm

There is no norm for the Turkish version, and Gate 4 of ADR-029 ($N \geq 1000$) is open. In order nevertheless to be able to show "where the person stands", the report uses a comparison distribution. Up to version 1.1 this distribution was a reference group computed on the fly from the scores of those who had completed the test in Rehnüma (ADR-032 §2). ADR-082 changed the source: the group was small (683 results at the time of the decision), had not been selected at random and contained trial completions. Personality percentiles today come from the scale's own published open data: Johnson's (2014) IPIP-NEO-120 data file (619,150 records; https://osf.io/wxvth/, `IPIP120.dat`). `Pusula-API/app/Profile/PublishedNorms.php` returns three values for each of the person's domain and facet scores: the percentile rank and the mean and standard deviation of the norm sample. The calculation does not consult the database; it is independent of how many people have taken the test in Rehnüma and of their scores.

The norm table is in the repository (`Pusula-API/database/data/norms_ipip_neo120.json`) and was produced from the open data by a script (`database/data/tools/build_ipip_norms.py`; the 95 MB data file has not been placed in the repository). The structure of the table follows the two stages of the form. Because stage 1 asks only the 60 core items, the published 120-item means do not fit it; two separate tables were therefore computed from the answers of the same people: the 60 core items for stage 1 and the 120 items for stage 2. The items were mapped to the official item key through the text of their English originals; 117 of the 120 items are original items and enter the norm, and the three rewritten O6 items have no counterpart in the original data (ADR-082 §3). At each stage the mean, the standard deviation and the percentile of every possible total score are stored for the five domains and 29 facets: a scale of k items has $4k + 1$ possible totals (49 values in a 12-item domain, 97 in a 24-item one, 9 in a two-item facet, 17 in a four-item one). The percentile is computed by the mid-rank (midrank) method: percentile = (number below + number equal / 2) / N × 100. Counting half of the

ties prevents the many ties in the discrete total scores of the scale from shifting the percentile systematically upwards or downwards. The table is the observed distribution itself; no normal distribution is assumed (ADR-082 §2).

Note. The `n` field in the table is 619,150 and is the number of records in the data file (`PublishedNormsTest` checks this value). The distribution of each scale was computed from the records that gave a valid answer to all the items of that scale; the number per scale is therefore smaller: in the domains 582,921–586,192 for stage 1 and 551,341–562,231 for stage 2; in the facets 604,190–616,867 (count from the table).

The person's score is looked up in the table (`PublishedNorms::place`). The 0–100 score is converted to its position over the possible totals of the scale; if the score falls between two table values, linear interpolation is used, and the result is rounded to an integer and kept in the range 1–99, that is, the report does not write 0% or 100%. Which stage's table is to be used is read from the item count: if all of the person's facets have been measured with four items, the domain scores are compared with the stage 2 table, otherwise with the stage 1 table; each facet likewise consults the table matching its own item count (2 or 4). In this way a 60-item domain score is not compared with the 120-item distribution. The tests check that the table carries the two stages and the five domains, that the percentile increases with the score and that four-item facets use the second-stage norm (`Pusula-API/tests/Feature/PublishedNormsTest.php`: `test_norm_tablosu_iki_asamayi_ve_bes_alani_tasir`, `test_yuzdelik_puanla_artar_ve_veritabanina_bakmaz`, `test_dort_maddelik_alt_boyutlar_ikinci_asama_normunu_kullanir`).

There are two exceptions. The first is the O6 facet: because three of its four items were rewritten (3.4.2), it has no counterpart in the original data; the build script does not write a norm for a scale fewer than half of whose items are original or which has a single original item left. There is no record for O6 in the table, and the function returns `null` for this facet; no percentile is given for O6. The Openness domain, by contrast, is normed with the remaining original items: 11 of the 12 items in stage 1 and 21 of the 24 items in stage 2 (ADR-082 §4). There is an approximation here: the person's O score is computed from the 12 or 24 items including the rewritten ones, whereas the distribution in the table is from the 11 or 21 original items; the comparison is made on the 0–100 equivalents of the two scales (inference from the code). The second is the interest areas: because there is no published norm table for the O*NET Interest Profiler, no percentile is given for the interest areas; the `interests` field of the function returns empty and the report shows only the 0–100 score (ADR-082 §5). The Rehnüma back end passes only the domain level of the norm to the summary; the facet percentiles remain in the full report (`Back-end/app/Services/Pusula/PersonalityReport.php`).

The language of the report is tied to this source: the text says that the personality percentiles are relative to the published international sample of the IPIP-NEO-120 (Johnson, 2014; 619,150 people); the source is named in an affirmative sentence, and negative expressions of the kind "not a Turkey norm" in the product text have been removed in the five languages (ADR-082 §6). The place where the limits are written is this monograph and the decision record. This is a reversal of the earlier decision: ADR-032 had rejected giving percentiles with Johnson's (2014) English norms on account of the difference of language and culture and of the Turkish version not having been validated, and had chosen the sentence "higher than P% of the N people who have completed it in Rehnüma". ADR-082 found a large, documented sample that anyone can download more defensible than a small and selected group; because the table and the build script are in the repository, the result is reproducible. The oth-

er options rejected were removing the percentile altogether, computing it with a normal distribution from the published mean and standard deviation (does not fit the 60-item stage, misleads at the extremes) and showing the norm only in the second stage.

The old mechanism remains in the code but is not connected to the report (ADR-082, Impact). `Pusula-API/app/Profile/ReferenceGroup.php` builds the domain, facet and interest distributions from the results completed with the same scoring version, computes the percentile with the same mid-rank formula and returns `null` if the group is below `norms_min_n` (default 30); its test is also in place (`ReferenceGroupTest`). The threshold of 30 people was far below the 200 that the EFPA considers "inadequate" for classical norming (EFPA 4.2.6; Evers et al., 2013); the group was therefore presented not as a norm but as an explicitly named comparison community. The class has been left in place to be reused with the Turkey data to be collected at Gate 4.

The limits of the published norm are written in ADR-082 and should be recorded here as well. First, the norm is an international sample consisting of volunteers who filled in the English form online; it is not a Turkey norm, and carrying it over to the Turkish adaptation is an assumption. The psychometric validation of the Turkish version has not been carried out, and 34 items were rewritten in the translation review (3.4); the percentile rests on the assumption that the Turkish items produce the same distribution as their English originals. Gate 4 ($N \geq 1000$, Turkey norm) therefore remains open. Second, the sample contains all ages and countries together; no separate norm by age and gender has been used. Third, the means of the norm sample are above the scale midpoint (53–67 in the domains): a raw score of 50 falls below the median in all five domains, and the same score may receive a lower percentile than under the previous reference group. Fourth, O6 has no percentile, and the norm of the O domain carries the approximation described above. Finally, although the source of the percentile has changed, the bands (3.7.1), the cut at 50 for the motif letters (3.8.1) and the scaling of the archetype prototypes (3.8.3) have remained on the raw scale; re-establishing these against the norm is a separate decision (inference).

Chapter references

- Bartram, D. (2005). The Great Eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*(6), 1185–1203. https://pubmed.ncbi.nlm.nih.gov/16316273
- Chauliac, M., Willems, J., Gijbels, D., & Donche, V. (2023). The prevalence of careless response behaviour and its consequences on data quality in self-report questionnaires on student learning. *Frontiers in Education, 8*, 1197324. https://doi.org/10.3389/feduc.2023.1197324
- Curran, P. G., & Hauser, K. A. (2019). I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items. *Journal of Research in Personality, 82*, 103849. https://www.gwern.net/doc/psychology/personality/2019-curran.pdf
- Doğan, Gündeğer, & Yurtçu. (2024). [Measurement invariance study with four different midpoint labels on the same 14 items]. https://acikerisim.aksaray.edu.tr/items/66f7f62d-cc6d-4fa5-95e9-fd7cbfbc20a6/full (reference to be verified)
- EFPA. (2013). *EFPA review model for the description and evaluation of psychological and educational tests* (Version 4.2.6). https://cms.bps.org.uk/sites/default/files/2022-07/110c_-_efpa_boa_test_review_model_version_4_2_6_approved.pdf

- Eggerth, D. E., & Andrew, M. E. (2006). Modifying the C index for use with Holland codes of unequal length. *Journal of Career Assessment, 14*(2), 267–275. https://doi.org/10.1177/1069072705283976
- Evers, A., et al. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema, 25*(3), 283–291. https://doi.org/10.7334/psicothema2013.97
- Gerlach, M., Farb, B., Revelle, W., & Amaral, L. A. N. (2018). A robust data-driven approach identifies four personality types across four large data sets. *Nature Human Behaviour, 2*, 735–742. https://doi.org/10.1038/s41562-018-0419-z
- Gerlach, M., Revelle, W., & Amaral, L. A. N. (2019). Reply to: Four personality types may be neither robust nor exhaustive. *Nature Human Behaviour, 3*(10), 1047–1048. https://doi.org/10.1038/s41562-019-0722-3
- Glavin, K. W., & Savickas, M. L. (2011). Interpreting Self-Directed Search profiles: Validity of the "Rule of Eight". *Journal of Vocational Behavior, 79*(2), 414–418. https://doi.org/10.1016/j.jvb.2011.01.005
- Hoeglund, T. J., & Hansen, J.-I. C. (1999). Holland-style measures of congruence: Are complex indices more effective predictors of satisfaction? *Journal of Vocational Behavior, 54*(3), 471–482. https://doi.org/10.1006/jvbe.1998.1675
- Hoerger, M. (2010). Participant dropout as a function of survey length in Internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior, and Social Networking, 13*(6), 697–700. https://doi.org/10.1089/cyber.2009.0445
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8
- Iachan, R. (1984). A measure of agreement for use with the Holland classification system. *Journal of Vocational Behavior, 24*(2), 133–141. https://doi.org/10.1016/0001-8791(84)90001-0
- International Test Commission. (2017). The ITC guidelines for translating and adapting tests (2nd ed.). *International Journal of Testing, 18*(2), 101–134. https://doi.org/10.1080/15305058.2017.1398166
- Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003
- Johnson, J. A. (2020). [Calibrating personality self-report scores to acquaintance ratings]. https://bpb-us-e1.wpmucdn.com/sites.psu.edu/dist/1/163537/files/2023/07/Calibrating2020.pdf (reference to be verified)
- Kajonius, P. J., & Johnson, J. A. (2019). Assessing the structure of the Five Factor Model of Personality (IPIP-NEO-120) in the public domain. *Europe's Journal of Psychology, 15*(2), 260–275. https://pmc.ncbi.nlm.nih.gov/articles/PMC7871748/
- Katahira, K., Kunisato, Y., Yamashita, Y., & Suzuki, S. (2020). Commentary: A robust data-driven approach identifies four personality types across four large data sets. *Frontiers in Big Data, 3*, 8. https://doi.org/10.3389/fdata.2020.00008
- Keirsey, D. (n.d.). *Temperament overview*. https://keirsey.com/temperament-overview/

- Kung, F. Y. H., Kwok, N., & Brown, D. J. (2018). Are attention check questions a threat to scale validity? *Applied Psychology, 67*(2), 264–283. https://doi.org/10.1111/apps.12108
- Lundmark, S., Silber, H., Rossmann, J., & Gummer, T. (2023). *Comparing types of attentiveness measures*. University of Gothenburg. https://www.gu.se/sites/default/files/2023-09/Comparing%20types%20of%20attentiveness%20measures%20-%20Lundmark%20Silber%20Rossman%20%20Gummer%20%282023%29.pdf (reference to be verified)
- Mavletova, A., & Couper, M. P. (2014). [Breakoff rates in mobile and desktop web surveys; GOR 2014 conference presentation]. https://cts.hse.ru/data/2014/04/03/1317876257/Mavletova_Couper_GOR_2014.pdf (reference to be verified)
- McCloy, R., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999b). *Development of the O*NET computerized Work Importance Profiler*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/DevCWIP.pdf
- McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality, 57*(1), 17–40. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085
- National Center for O*NET Development. (n.d.-a). *O*NET Interest Profiler Short Form psychometric characteristics*. https://www.onetcenter.org/dl_files/IPSF_Psychometric.pdf
- National Center for O*NET Development. (n.d.-b). *O*NET Interest Profiler manual*. https://www.onetcenter.org/dl_files/IP_Manual.pdf
- National Center for O*NET Development. (n.d.-c). *Work Importance Locator: Instructions*. https://www.onetcenter.org/dl_files/WIL_zips/WIL-Instr-deskv.pdf
- National Center for O*NET Development. (n.d.-d). *O*NET Tools Developer License*. https://www.onetcenter.org/license_toolsdev.html
- Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology, 22*(3), 453–460. https://doi.org/10.1016/j.jcps.2011.08.002
- Pittenger, D. J. (1993). The utility of the Myers-Briggs Type Indicator. *Review of Educational Research, 63*(4), 467–488. https://doi.org/10.3102/00346543063004467
- Primi, R., De Fruyt, F., Santos, D., Antonoplis, S., & John, O. P. (2020). True or false? Keying direction and acquiescence influence the validity of socio-emotional skills items in predicting high school achievement. *International Journal of Testing, 20*(2), 97–121. https://doi.org/10.1080/15305058.2019.1673398
- Putka, D. J., Kell, H. J., Voss, N., Oswald, F. L., & Lewis, P. (2024). *Revisiting the O*NET Work Styles domain*. National Center for O*NET Development. https://onetcenter.org/dl_files/Work_Styles_New.pdf
- Rammstedt, B., Kemper, C. J., & Borg, I. (2013). Correcting Big Five personality measurements for acquiescence: An 18-country cross-cultural study. *European Journal of Personality, 27*(1), 71–81. https://doi.org/10.1002/per.1894
- Revilla, M., & Höhne, J. K. (2020). How long do respondents think online surveys should be? New evidence from two online panels in Germany. *International Journal of Market Research, 62*(5), 538–545. https://www.sowi.uni-mannheim.de/media/Einrichtungen/gip/Bilder/Dokumente/revilla_hoehne__2020__-_online_survey_length.pdf

- Rounds, J., Wee, C. J. M., Cao, M., Song, C., & Lewis, P. (2016). *Development of an O*NET Mini Interest Profiler (Mini-IP) for mobile devices: Psychometric characteristics*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/Mini-IP.pdf
- Somer, O., Korkmaz, M., & Tatar, A. (2002). Beş Faktör Kişilik Envanteri'nin geliştirilmesi-I: Ölçek ve alt ölçeklerin oluşturulması. *Türk Psikoloji Dergisi, 17*(50), 21–36. https://avesis.mcbu.edu.tr/yayin/9f580945-7e2e-48e8-9a28-d68cfc6c173e/development-of-five-factor-personality-inventory-bes-faktor-kisilik-envanterinin-gelistirilmesi-i-olcek-ve-alt-olceklerin-olusturulmasi
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of Big Five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718–737. https://doi.org/10.1037/0022-3514.94.4.718
- Süerdem, A., & Erkök, B. (2016). Assessing the reliability and validity of a shorter version of RIASEC in Turkish. *SHS Web of Conferences, 26*, 01063. https://doi.org/10.1051/shsconf/20162601063
- Welz, M., & Alfons, A. (2023). *When respondents don't care anymore: Identifying the onset of careless responding* [Preprint]. arXiv. https://arxiv.org/abs/2303.07167

4. Validation Protocol and Simulation Findings

This chapter describes how the reliability and validity of the measurement design presented in Chapter 3 have been, and will be, tested. It first summarises the validity and reliability framework on which the work rests; it then presents the five-gate validation protocol made binding by ADR-029. The centre of gravity of the chapter is Gate 0, the only gate of the protocol passed so far: the method of the simulation run with synthetic persons against the real test service, its findings, the decisions that were carried into the code on the basis of those findings, and the limits of the simulation. The chapter closes with the analysis plan for Gates 1–4, which require real user data, and with how the language of the report is tied to the gates.

The main thesis of the chapter is this: Gate 0 has shown that *the scoring code* works correctly; it has not shown, and cannot show, that the measurement is valid in Turkish. The real contribution of the simulation came from an unexpected place: it quantified how fragile the categorical labels (the four-letter code, the first Holland letter, the archetype) are even when the latent trait does not change at all, and it caught two concrete design flaws before the very first user (ADR-029 Rationale).

4.1 Validity and reliability framework

4.1.1 Standards and guidelines

The overarching framework of the work is the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014). The first part of the Standards consists of three foundational chapters: validity; reliability/precision and errors of measurement; fairness in testing. In the 2014 edition fairness was made a foundational principle on a par with validity and reliability. The Standards are not a legal requirement but criteria and guidance for sound testing practice (`dogrulama_protokolu.md` §9). For Rehnüma the Standards also carry legal weight: §3(b) of the O*NET Tools Developer License requires, for every product whose items have been modified (translation included), a validation study in accordance with the Standards before release (National Center for O*NET Development, n.d.-d). The protocol described in this chapter is therefore not only a scientific choice but also a condition of the licence.

The view that validity is not a single coefficient but an accumulation of evidence supporting an interpretation of scores is the basis of the validity chapter of the Standards. Messick's unified concept of validity is cited in the literature as the theoretical source of this view; however, since the references to Messick's works are not included in the research notes in the repository, it is mentioned here only as a framework and has not been included in the chapter references. The practical consequence for the work is clear: for Rehnüma the question asked is not "is the test valid" but "which score interpretation is supported by which evidence", and each gate collects the evidence for a particular kind of interpretation (4.2).

For the adaptation itself the reference text is the ITC guidelines for translating and adapting tests (ITC, 2017). The second edition consists of 18 guidelines divided into six sections: pre-conditions (intellectual property permission, overlap of the construct), test development (an appropriate transla-

tion design, evidence for the equivalence of the items and the rating scale, a pilot), confirmation (an adequate sample, statistical evidence of equivalence, norms, reliability and validity), administration, score scales and interpretation, and documentation. The sample recommendations of the guidelines determine most of the numbers in the protocol: "at least a modest sample (e.g., 100)" for pilot item analysis, "perhaps 300 or more" for factor structure, at least 200 per version for DIF, and at least 500 for IRT. The documentation guideline (Doc-1) requires the technical document to be detailed enough for someone else to replicate the work and to answer five questions: what evidence is there for the utility of the construct in the new population, which item data were collected from which sample, what other validity data are there, how were the data analysed, and what are the results (`dogrulama_protokolu.md` §1). This monograph is itself an attempt to answer the part of these five questions that can be answered today.

The third reference text is the EFPA test review model (EFPA, 2013; Evers et al., 2013). The model grades norms, reliability, construct and criterion validity, and computer-generated reports separately from 0 (no information) to 4 (excellent), and it is the numerical basis of most of the thresholds in the protocol.

4.1.2 Reliability thresholds and tiers of use

The reliability threshold depends on the purpose of use. EFPA 4.2.6 grades internal consistency as inadequate below .70, adequate at .70–.80, good at .80–.90 and excellent at ≥ .90; it grades test–retest as inadequate below .60, adequate at .60–.70, good at .70–.80 and excellent at ≥ .80; and for the sample of a reliability study it counts < 100 as inadequate, 100–200 as adequate and > 200 as large. The origin of the common ".70 is adequate" rule is contested: Lance, Butts, and Michels (2006) show that Nunnally recommended .70 only for the early stage of research and .90 for applied decisions. Johnson's (2014) judgement on the IPIP-NEO-120 rests on the same distinction: the facet scales are adequate for research and probably inadequate for important decisions about individuals.

Rehnüma's "open mode" is a low-stakes use: the person receives the information for their own interest, and no selection or placement decision is made on the basis of the result (as cited from the ITC guideline on computer-based testing; research report, "Puanlama, kalite ve rapor" [Scoring, quality and report]). Nevertheless, because the report gives individual feedback, the thresholds have been kept above the research level. The tiers of use that the research note proposes by inference are given in Table 4.1.

Table 4.1. Report language permitted by reliability level (tiers of use).

Tier	Facet (4 items)	Domain (24 items)	Permitted language	Status
Preliminary screening	α/ω ≥ .60	α/ω ≥ .80	Raw score, within-person ranking, "first impression", facet "tendency"	Target (Gates 1–2)
Individual feedback	α/ω ≥ .70	α/ω ≥ .85	Banded facet interpretation, domain score with confidence interval	Target (Gate 3)
Definite judgement	—	≥ .90	A "you are X" judgement based on a single facet score	At no gate

Note. The tiers are an inference of `dogrulama_protokolu.md` §2 and do not rest on a normative source; they are consistent with the EFPA (2013) grades and with Johnson's (2014) caution. To date none of these numbers has been measured for the Turkish version.

4.1.3 Types of evidence

The protocol distinguishes five types of evidence and ties each to a separate gate: the correctness of the scoring code (Gate 0), internal consistency at item level and the absence of keying errors (Gate 1), internal structure, that is, the replication of the five factors and of the RIASEC circular order in Turkish data (Gate 2), temporal stability and relations with external criteria (Gate 3), and norms and fairness across groups (Gate 4). This ordering follows the flow of the ITC and rests on the principle that the claim of a gate can only be built on the evidence of the preceding gates: the factor structure of a scale with a keying error, and the norm of a scale with a broken structure, are meaningless.

4.2 The five-gate protocol

ADR-029 §1 makes the protocol binding: each gate is a condition, a threshold and an output; the language of the report changes with the gate, and the claim of a gate that has not been passed does not enter the report. Table 4.2 gives the gates as they are written in the ADR.

Table 4.2. The five validation gates: condition, decision criterion and status (ADR-029 §1).

Gate	Condition	Threshold (decision criterion)	Status
0 Simulation	≥ 300 synthetic persons, real API	Facet recovery r ≥ .80 (stage 2), domain ≥ .85; careless-responder detection sensitivity ≥ .95, false alarm ≤ .05	Passed (23 October 2026)
1 Pilot	N ≈ 100 real users	Stage 2 facet α ≥ .60, domain α ≥ .80; item–total r ≥ .20; distribution of the acquiescence index	Pending
2 Structure	N ≥ 300	EFA 5 factors, Tucker congruence ≥ .85; RIASEC hexagonal order (randomisation test p < .05)	Pending
3 Stability	N ≥ 60, 4–8 weeks	Domain retest r ≥ .80; same motif ≥ 60% (≥ 85% excluding balanced letters)	Pending
4 Norms	N ≥ 1000	Percentiles banded by age/gender; Barnum pruning (a sentence with `report_feedback` "doesn't fit" ≥ 40% is removed)	Pending

Note. The table is reproduced from ADR-029. The validation protocol report and `dogrulama_protokolu.md` §10 propose stricter thresholds for some gates; the differences are shown in Table 4.3. Under the working rules it is the ADR that is binding.

Each gate opens one type of interpretation. The claim opened by Gate 0 is "the scoring is correct", and with this claim the product may use only raw scores and the "first impression" language. Gate 1 opens the claim of a pilot version at item level and the "tendency" wording of the facet bands; Gate 2 opens the removal of the "preliminary" label from the archetype and the code, and the provisional percentile; Gate 3 opens the individual report with confidence intervals; and Gate 4 opens the "Turkish norm" claim (validation protocol report, gate table). This mapping is taken up together with the language of the report in 4.8; the source of the percentile changed on 22 November 2026 with ADR-082, and that change is also described there.

4.2.1 Threshold differences between documents

The protocol exists in two written forms: the thresholds proposed by the research note and the report, and the thresholds adopted by ADR-029. The two coincide at many points, but at four gates the ADR has chosen a looser or a different criterion (Table 4.3).

Table 4.3. Differences in gate thresholds between the decision document and the research report.

Gate	Criterion	ADR-029	Research report / note	Used in the code
0	Domain recovery	r ≥ .85	r ≥ .90	— (the script does not check thresholds)
0	Facet recovery (2 items)	no threshold	r ≥ .70	—
0	Careless-response detection	sensitivity ≥ .95, false alarm ≤ .05	no numerical threshold ("sensitivity reported")	—
1	Item–total correlation	r ≥ .20	r ≥ .30 (flag < .20)	`psychometrics.py`: weak item < .30
1–2	Gate for facet and domain internal consistency	Gate 1: stage 2 facet α ≥ .60, domain α ≥ .80	Gate 2: facet α/ω ≥ .60, domain ≥ .80	The script reports α on every run; it makes no gate distinction
2	Tucker congruence	≥ .85	≥ .90	`psychometrics.py` comment: ≥ .85 acceptable, ≥ .95 equal
2	RIASEC circular order	p < .05	p < .05 and CI ≥ .60; CUS ≥ 80%	The script reports CI and p
3	Retest sample and interval	N ≥ 60, 4–8 weeks	N ≥ 100, 2–4 weeks	`simulate.py` comment: "4–8 weeks"
3	Facet retest	no threshold	r ≥ .70	—

Note. Sources: ADR-029 §1; `docs/research/reports/Kişilik testi doğrulama protokolü.md` (gate table); `dogrulama_protokolu.md` §10; `docs/research/scripts/psychometrics.py`, `simulate.py`.

Some of the differences are reasoned. For domain recovery, the .90 threshold lies above a ceiling that the synthetic person model itself does not allow to be reached (4.4.1); the ADR's .85 is consistent with this observation (inference). Lowering the retest sample to 60, on the other hand, is a pragmatic decision that takes into account the data a small team can collect in the first months, but it should not be forgotten that EFPA counts fewer than 100 as "inadequate". In a retest of 60 persons the standard error of a rate around 60% is about 6.3 points and the 95% confidence interval about ±12 points (calculation); the "same motif ≥ 60%" criterion of Gate 3 can therefore yield only a rough decision at N = 60. For the item–total threshold at Gate 1, by contrast, the code is stricter than the ADR: `psychometrics.py` places items below .30 on the weak list. This can be read as a deliberate distinction between a decision criterion (.20) and a review list (.30); however, no such distinction is written in the ADR.

Note. The gate table of the validation protocol report counts not only the simulation but also the TRAPD translation record as a condition for Gate 0. ADR-029 defines Gate 0 by the simulation alone and says "passed". The second independent translation, the reconciliation record and the cognitive interviews have not yet been carried out (3.4.4); by the report's definition Gate 0 has been half passed.

4.3 Gate 0 method

4.3.1 Purpose and template

Gate 0 is a parameter recovery study. The simulation template of Feinberg and Rubright (2016) prescribes generating responses from known latent traits, passing these responses through the product's own scoring code and measuring how well the returned scores recover the true values; the outcome criteria include correlation, bias and mean squared error, and the random seed and the number of

replications are reported. The code is verified "by reproducing known results". This study answers a single question: do the scoring code, the reverse-keyed items, the stage logic, the quality indicators and the label rules work correctly? It says nothing about whether the translation is good; the research note particularly stresses that the two questions must be kept apart in the report (`dogrulama_protokolu.md` §7).

The script that runs the simulation is `docs/research/scripts/simulate.py`. The script takes the test not through a mock or a unit test but over HTTP against the real Pusula-API running locally: for each synthetic person a test is started with `POST /assessments`, the answers of each section are sent with `PUT /assessments/{token}/sections/{key}`, `POST /assessments/{token}/complete` is called at the end of the stage, and the returned result is stored. The verification thus covers the whole of the endpoints the client will see, the server-side validator, the scorers, the profile builder and the occupation matcher. So that the thousands of requests of the simulation do not hit the service's per-minute request limit, the limit has been made configurable (`PUSULA_RATE_PER_MINUTE`; default 600, raised locally; ADR-029 §5).

One implementation detail reflects a property of the design. Because the scoring keys are not sent to the client (3.2.2), the script cannot learn from the API which facet each item belongs to or whether it is reverse-keyed. The `hidden_form` function reads this information from the stored form in the local SQLite database. In this sense the simulator is an "oracle" that knows what the client does not; in order to answer according to its own latent facet score, the synthetic person has to know which item belongs to which facet.

The simulation was run with form v4 (ADR-026); all Gate 0 numbers in this chapter belong to that form. Form v5 (ADR-044), adopted on 6 November 2026, adds two ranking steps without changing the Likert bank of v4: `self_place` at the beginning of stage 1 and `rank` at the beginning of stage 2. These two steps were not part of the simulation; because full ranking is a separate within-person (ipsative) measurement, it has been placed beside the Likert items rather than in their place, the Five-Factor and interest scores remain as in v4, and the new steps are monitored under a separate heading in the Gate 1 report (ADR-044 Rationale). The script's `answer_section` function answers only the four section types of v4 (`likert`, `expressed`, `card_sort`, `choice`); since the service now serves form v5 (`FormBuilder`), the script cannot be rerun until these two section types are added (inference from the code).

4.3.2 Synthetic person model

Each synthetic person (the `Persona` class) carries five domains, thirty facets, six RIASEC areas, a value ranking and a rhythm preference. Table 4.4 gives the parameters of the model from the code.

Table 4.4. Parameters of the synthetic person and response model (`simulate.py`).

Component	Model	Parameter
Big Five domains	Multivariate normal, by Cholesky decomposition	Correlation between domains .00–.30 (e.g. E–O .30, S–C .25, S–E .20)
Facets	Facet = 0.7 × domain + 0.7 × specific component	Facet–domain correlation ≈ .71; between facets of the same domain ≈ .50 (calculated)
RIASEC	$r_i = 0.75\, b_i + 0.35\, b_{i+1} + 0.35\, b_{i-1}$ (hexagonal adjacency)	Adjacent ≈ .65, second neighbour ≈ .15, opposite area 0 (calculated)

Component	Model	Parameter
Likert response	y* = θ × key + acquiescence + N(0, 0.9); thresholds −1.5, −0.5, 0.5, 1.5	Discrimination 1 for all items
Individual acquiescence	N(0, 0.15)	Fixed for each person
Expressed interest	Top two and last by the order of latent RIASEC + N(0, 0.5)	—
Card sort	4 cards in each of 5 columns, by value order with N(0, 0.8) noise	Value order a random permutation
Rhythm	One of five options; the same option for all three questions	Probabilities .15, .30, .25, .20, .10
Page time	Section as a single page	Honest and straightliner 4.0 s/item, random 1.2 s/item

Note. The correlations marked "calculated" were derived analytically from the model equations (the authors' calculation); the script does not report them.

The response model is a linear-threshold model that resembles the graded response model but is simpler: the latent facet score θ is multiplied by the keying direction of the item (+1 or −1), the person's acquiescence tendency is added, a normally distributed error is added, and the result is cut into the categories 1–5 by four fixed thresholds. The template proposed by the research note (item discrimination log-normal, thresholds noisy from item to item) is richer (`dogrulama_protokolu.md` §7); the script assumes all items to be equivalent. For interest items the key is always +1. From the model equations, item reliability before categorisation can be calculated as about .55 (true variance 0.98, error variance 0.81; calculation).

4.3.3 Types of persons

The script generates three types of persons. The default careless rate is 10%; half of the careless persons are straightliners (3 to every item) and half are random responders (1–5 with equal probability to every item). An honest person gives the expected answer to the instructed-response items; careless persons do not. The straightliner also answers 3 to an instructed-response item; the random responder gives a random value to an instructed-response item and therefore passes each check by chance with probability 1/5. Table 4.5 summarises the types of persons and what the quality indicators are expected to produce for them.

Table 4.5. Types of synthetic persons and the expected quality outcome.

Type	Rate	Likert answer	Instructed-response item	Page time	Expected flags	Expected level
Honest	90%	Model response	Expected answer	4.0 s/item	None or a single flag	`ok` / `note`
Straightliner	5%	Always 3	3 (expected 2, 4, 2)	4.0 s/item	`longstring`, `check_failed`	`uninterpretable` (instruction + longstring)
Random	5%	1–5 with equal probability	Random	1.2 s/item	`too_fast`, most probably `check_failed`, `inconsistent`	`uninterpretable` (fast-page rate ≥ 99%)

Note. The expected levels are the application of the `Quality::evaluate` rules (3.6.3) to these answer patterns. For the straightliner the odd–even correlation is undefined (zero variance) and the `inconsistent` flag is not produced.

The limit of the simulation with respect to careless-response detection can be read directly from the table: both careless types are defined in such a way that they directly trigger one of the severe-case rules. This tests that the detection code works, but it does not test how *difficult* the cases are that the detection catches (4.6).

4.3.4 Flow, retest subsample and criteria

The run was carried out on 23 October 2026 on form v4 with 300 synthetic persons and 60 retests (ADR-029; validation protocol report). The seed is 42 (the script's default). Each person completes both stages: the stage 1 sections are sent and completed, then the stage 2 sections are sent and completed; in this way both the summary (stage 1) and the full report (stage 2) result are obtained for each person. For the retest, the first 60 honest persons take the test again from the beginning with *the same latent traits* but with new noise. Because a new test is created for the second sitting, the order of items within sections is also reshuffled. The script's comment describes this as "as if 4–8 weeks later; trait assumed constant". Because the latent trait does not change at all between the two administrations, this retest is the ceiling of a real retest.

The `evaluate` function produces four groups of criteria. Recovery: for honest persons only, and separately for stage 1 and stage 2, the Pearson correlation between the latent domain, facet and RIASEC values and the returned 0–100 scores; in addition, the rate at which the latent strongest RIASEC area is found in the returned three-letter code, and the rate at which the latent first value is found among the returned top two values. Quality: the rates of being flagged at the `review` or `uninterpretable` level among careless and among honest persons, and the distributions of levels. Stability: in the retest subsample, the correlation of domain and facet scores between the two administrations, together with the rates at which the four-letter code, the first archetype, the temperament, the first Holland letter and the first value remain the same (stage 2; for the code, also stage 1). Distribution: among honest persons, the frequencies of the first archetype, the temperament and the code, the mean number of balanced letters and the mean number of strengths; this last group checks whether there is any piling-up on a single type.

4.4 Gate 0 findings

4.4.1 Facet and domain recovery

The scoring side passed the gate clearly (Table 4.6). At stage 1, that is, with two items per facet, facet recovery has a median of r = .81 and a minimum of .78; domain recovery lies in the range .84–.86. At stage 2, with four items per facet, the facet median rises to .89, the minimum to .86, and the domain values to .87–.89. RIASEC recovery is .92–.94.

Table 4.6. Gate 0 recovery and retest correlations (honest synthetic persons).

Measure	Stage 1 (2 items per facet)	Stage 2 (4 items per facet)	Threshold (ADR-029 / report)
Facet recovery, median r	.81	.89	— / .70 (2 items), .80 (4 items)
Facet recovery, minimum r	.78	.86	≥ .80 (stage 2)
Domain recovery r	.84–.86	.87–.89	≥ .85 / ≥ .90

Measure	Stage 1 (2 items per facet)	Stage 2 (4 items per facet)	Threshold (ADR-029 / report)
RIASEC recovery r	—	.92–.94	—
Domain retest r	—	.91–.95	≥ .80 (Gate 3)
Facet retest r, median (minimum)	—	.79 (.71)	— / ≥ .70

Note. The run was carried out with form v4. The values are reproduced from ADR-029 §2 and the validation protocol report; the raw output files (`report.json`, `raw.json`) are not present in the repository. The sources do not state explicitly which stage the RIASEC value belongs to; it has been assumed to belong to the 60-item stage 2.

Three observations are important for interpreting the results. First, $r < .50$ did not occur in any facet. By the acceptance criterion of the research note, a facet below .50 is the sign of a keying error or a facet–domain assignment error; hence the keying direction and facet assignment of the 120 items, including the nine one-directional facets and the reversal of the S domain, are processed correctly. Second, facet recovery increased with the number of items in the expected direction (.81 → .89); this shows that the A/B/C blocks are combined correctly at stage 2.

Third, the domain values remained slightly below the .90 proposed by the report. It was considered that this is most probably not a scoring flaw but the ceiling of the generating model (validation protocol report). The model equations support this interpretation: the domain score rests on the mean of six facets, and because the facets are generated as $0.7 \times$ domain $+ 0.7 \times$ specific component, the correlation between the mean of six facets and the latent domain stays at about .93 even with no item error at all ($0.7 / \sqrt{(0.49 + 0.49/6)}$; calculation). Once item-level error and the 1–5 categorisation are added, .87–.89 is a reasonable value. The check proposed by the report is to compute the correlation of the domain score with the mean of the latent facet values instead of with the latent domain; if this value exceeds .90, the domain aggregation rule is clean. This additional computation has not yet been carried out.

4.4.2 Retest

In the retest subsample the domain correlations are .91–.95, the median of the facet correlations is .79 and the lowest is .71; the stability targets of the protocol (domain ≥ .80, facet ≥ .70) are met. These values should be read as the ceiling attainable in the real world: the latent trait of the synthetic person did not change between the two administrations, whereas in a real 4–8 week retest (the Gate 3 interval), state effects, mood and recall are added. A realistic external reference is the finding of $r = .86$ for domains and .80 for facets at a six-month interval in the Czech adaptation of the BFI-2 (validation protocol report).

4.4.3 Careless-response detection

Sensitivity was found to be 1.00 among careless synthetic persons and the false alarm rate 0 among honest ones (ADR-029 §2): all straightliners and random responders were flagged at least at the `review` level, and no honest person reached that level. The result meets the Gate 0 thresholds (sensitivity ≥ .95, false alarm ≤ .05). However, as shown in 4.3.3, this is an easy case: synthetic carelessness is either straightlining or fully random responding, and both directly trigger the severe-case rules. In the

literature, longstring catches straightlining but misses inconsistent carelessness such as "midpoint marking" (Welz & Alfons, 2023); sensitivity is expected to fall in real data, and the thresholds should be recalibrated at Gate 1.

For the acquiescence index the simulation provided two separate pieces of information. Among honest persons the index was at most 0.57, and among straightliners 2.0; the 1.0 threshold separates these two groups (ADR-030 Rationale). The first form of the index was related to the domain scores ($r = .33$ with E, $-.39$ with A); the form restricted to balanced facets gave correlations of E $-.04$, O $-.06$, C $-.07$, A $-.24$, S $-.26$ in a rerun of 98 persons (ADR-030 §2; 3.5.2).

Note. On this last point there is a difference between the documents and the code that needs to be explained. By the definition of `Quality::acquiescence`, the index of a person who answers 3 to every item should be 0; the statement in ADR-030, "2.0 among straightliners", is not compatible with this definition. The statement corresponds to a profile that gives the most extreme "agree" answer (5) to every item ($5 - 3 = 2.0$); the test `ScoringTest::test_kabul_etme_indeksi_her_seye_katilani_isaretler` also tests a profile of this kind. The straightliners in `simulate.py`, however, give 3 to every item. Taking the code as the basis, the value 2.0 cannot be produced by the straightliners of `simulate.py` who answer 3; the value is thought to belong to a profile that "says 5 to everything" or to a run of the script with different settings (inference). The relevant sentence of ADR-030 should be clarified in this respect. The distinction matters: a person who straightlines on the middle option is caught not by the acquiescence index but by longstring and the instructed-response item.

4.4.4 Label stability

The main finding is in the label layer. Even with the latent trait held constant between the two administrations, the categorical labels changed at a marked rate (Table 4.7).

Table 4.7. Stability of categorical labels in the Gate 0 retest (N = 60).

Label	Rate remaining the same	Additional observation	95% CI, approximate (calculated)
Four-letter mind motif	57%	Clear letters never flipped; 36% of balanced letters flipped	±12.5 points
First letter of the Holland code	45%	First letter of the second administration is one of the top two areas of the first: 75%	±12.6 points
First archetype	70%	77% among those with a difference of ≥ 10 points between the top two archetypes	±11.6 points
Temperament	80%	More stable than the code because it is derived from two letters	±10.1 points
First work value	73%	Rate at which the first value remains in the top two: 99%	±11.2 points

Note. The rates are reproduced from ADR-029 §2 and the validation protocol report. The confidence intervals were calculated by the authors with the normal approximation as $\sqrt{p(1 - p)/60} \times 1.96$. The breakdowns "flip rate of balanced letters", "one of the top two" and "difference ≥ 10 points" are not defined in the `evaluate` function of `simulate.py`; it appears that they were computed by additional analysis on the raw output (inference). Moreover, the `values_top1_in_top2_rate` criterion in the script is not a retest criterion but the rate at which the latent first value is found in the returned top two (recovery); it is not clear in the sources whether the value "99%" belongs to this criterion or to the retest.

The meaning of these numbers is clear: the variability comes not from the scale but from the cut-point. Converting a continuous score into a category moves the label of persons near the boundary by as much as the measurement error, even when the true trait does not change (MacCallum et al., 2002). The fact that clear letters never flipped and that the flips were concentrated entirely in balanced letters shows that the rule works in the right place and that the problem is not a code error but

the structural cost of dichotomisation. The reflections of the same mechanism in real data are known: a type change of 39–76% within five weeks on the MBTI (as reported by Pittenger, 1993) and first-letter agreement between the Mini-IP and the IP-SF remaining at $\kappa = .73$ (Gregory & Lewis, 2016).

That the first Holland letter stays at 45% is far below the 75% target the protocol proposes for the same first letter. According to the report's interpretation, an important reason for this is that most of the synthetic RIASEC profiles are undifferentiated: the difference between the top two of the six areas generated with hexagonal adjacency is small for most persons, and when the difference between the top two areas is small the first letter is "a coin toss". For this reason the same rate needs to be reported separately for differentiated profiles. Value stability (73%) is higher than in the real data of the WIP-C (first value again in first place 62–70%, in the top two 88%; McCloy et al., 1999b); this is the typical sign of synthetic superiority.

The distribution criteria (frequencies of archetype, temperament and code, mean number of balanced letters and of strengths) are computed in the script; however, their values are not reported in the ADR or in the report. Whether there is piling-up on a single archetype or temperament therefore cannot be reported in this monograph.

4.5 Decisions taken on the basis of the findings

The value of Gate 0 lies in the findings returning to the code. ADR-029 §2 paired each finding with a decision; Table 4.8 gives this mapping together with the code location where the decision was implemented.

Table 4.8. Gate 0 findings, the decisions taken and where they were implemented in the code.

Finding	Decision	Code location
Stage 1 facet r median .81, minimum .78	Stage 1 facet rows continue to be marked as "first impression"	`ProfileBuilder::build(provisional)`, `facetSides(certain)`
Stage 2 facet r median .89, minimum .86	Facet texts are opened at stage 2	Stage logic of the report sections (Rehnüma)
Careless-responder detection sensitivity 1.00, false alarm 0	The existing indicators remain; the acquiescence index is added	`Quality::evaluate`, `Quality::acquiescence`
Same motif 57%; 36% of balanced letters flipped	Balanced threshold 6 → 8 points; if there is a balanced letter, neighbouring motifs are shown as alternatives	`ProfileBuilder::BALANCED`, `ProfileBuilder::alternatives`
First Holland letter 45%; one of the top two 75%	If the difference between the top two areas is ≤ 5 points, `top_two_close`; the report says "neck-and-neck"	`ExpressedInterest::apply`
First archetype 70% (77% at difference ≥ 10)	Percentages continue to be shown, and the total is completed to exactly 100; the prototypes will be rebuilt by clustering with Gate 2 data	`ProfileBuilder::archetypes`
Temperament 80%, first value 73%	Values are presented as "the two that stand out"	`ValuesScorer(top)`

Note. The decision column is taken from ADR-029 §2, the code locations from the authors' code review.

4.5.1 The balanced threshold and alternative motifs

The most concrete decision is the raising of the balanced threshold from 6 points to 8 points. The rationale is that in the simulation the standard deviation of the retest difference was found to be about 6–7 points: when the score of a letter is closer to 50 than 6 points, the typical difference between two administrations can carry it to the other side of the boundary. Raising the threshold to 8 widens the balanced band and further reduces the probability that letters labelled "clear" will flip. The cost is that more people see at least one balanced letter. The validation protocol report characterises this not as an error but as honesty.

The threshold change does not by itself increase stability; what changes is how instability is presented to the user. Showing the neighbouring motifs formed by flipping the balanced letters as "alternatives" in the report (3.8.1) ensures that, if a balanced letter flips at retest, the new motif the person encounters is already mentioned by name in the report. The criterion of Gate 3 reflects this design: the rate at which the motif remains the same must be ≥ 60% in the whole sample and ≥ 85% when balanced letters are excluded (ADR-029 §1).

Note. The validation protocol report records that the effect of the corrections should be reported by rerunning with the same seed and that the question "to what did the percentage of identical codes rise at threshold 8" has not yet been answered. The code change (`BALANCED = 8.0`) has been made, but the result of the rerun is not in the repository. The 57% value in Table 4.7 belongs to the 6-point threshold.

4.5.2 The "neck-and-neck" flag

The second decision concerns the Holland code. If the difference between the top two areas is 5 points or less, the `top_two_close` flag is raised and the report says that the first letter is "neck-and-neck". The threshold was chosen as the counterpart on this scale of the "rule of 8" of the SDS; in Glavin and Savickas's (2011) data on 2,397 students, a code difference above the measurement error raised the hit rate of the choice of major from 40% to 55% (validation protocol report). The code comment states that the decision rests directly on this finding (`ExpressedInterest.php`: "in the simulation first-letter stability was 45%, while 'one of the top two' was 75%"). The research report recommended that the 5-point threshold be calibrated empirically for our own scale: at Gate 3 it should be measured at which difference threshold the first letter remains the same 75% of the time at retest.

4.5.3 The acquiescence index and the completion of percentages

The third decision is the addition of the acquiescence index and its restriction to balanced facets (3.5.2; ADR-029 §4, ADR-030 §2). The simulation was used twice here: it showed the relation of the first version to the domain scores, and it confirmed that the corrected version brought this relation below the .30 threshold. The fourth decision removed a small but visible flaw: when the archetype percentages were rounded independently, the total could be 99 or 101; completing it to exactly 100 by the largest remainder method removed this inconsistency.

4.5.4 Decisions not taken

The simulation also supported *not* taking certain decisions. Removing the code or the archetype from the report was rejected; the user decision (ADR-028) calls for a visible label, and uncertainty is preserved not by being hidden but by being shown. A score correction for acquiescence (ipsatisation) was rejected; without norms a correction would be a new and arbitrary threshold. Direct use of the

Gerlach centres was rejected; the numerical centres could not be verified from the primary source. Finally, not publishing the test until validity evidence arrives was rejected: Gates 1–4 require real users, and the product collects data while remaining honest through the language of "first impression" and "technical appendix" (ADR-029 Rejected alternatives).

4.6 Limits of the simulation

All the numbers given by the simulation are *upper bounds* and are expected to fall in real data for the following reasons.

The model shares the assumptions of the scoring. The synthetic persons were generated from the structure that the scoring code assumes: each item measures only its own facet, the discriminations are equal, the thresholds are symmetric, and there are no cross-loadings. Real items, by contrast, have translation-induced shifts of meaning (3.4.2), the source's own weak items (the item–total correlations of .21–.33 reported by Johnson) and social desirability. For this reason the simulation cannot see a translation error; for example, if the item "Sorumluluğu üstlenirim" (literally "I take on responsibility"; English original "Take charge") were measuring C content, the simulation would not notice it, because the synthetic person answers this item according to the latent E3 score. The deliberate error injection proposed by the research note (flipping the key of an item, swapping two facets) has also not yet been carried out (`dogrulama_protokolu.md` §7).

The carelessness model is easy. The synthetic careless persons either answer 3 to every item or answer fully at random and very fast. Real carelessness is often partial (it increases towards the end of the test), takes the form of midpoint marking, or is confined to one section. These types are absent from the simulation; that sensitivity came out at 1.00 therefore shows that the code works, not the real power of the detection. The flag rate expected in the literature is 5–15%; for a rate exceeding 20% the cause should first be sought in the interface and in item length (`dogrulama_protokolu.md` §6).

The strength of acquiescence is small. The individual acquiescence tendency of the synthetic persons was generated with N(0, 0.15). The wide variability in adolescent data shown by Soto et al. (2008) is not represented in this model. The simulation therefore does not estimate how much profile distortion the keying imbalance will produce in adolescents; this question has been left to Gate 1 (ADR-030 §3).

Timing is a single page. The script sends each section as a single "page" and writes the page time as the product of the number of items and a fixed speed. In the real interface each section is divided into pages of six items (3.1.3) and the speed indicator rests on the page rate; page-level variability is absent from the simulation.

Some sub-layers were tested trivially. Because the synthetic person answers the three rhythm questions with the same option, the "match rate" of the rhythm type is high by definition; this measure tests the aggregation rule of the code, not the chronotype measurement. In the card sort, noise was added only to the value order; the real instability arising from the ipsative structure of the WIL is not represented, and the fact that value stability came out higher than in real data is the sign of this (4.4.4).

The retest is stationary and small. The latent trait did not change at all between the two administrations; real stability will remain below this. In the subsample of 60 persons the 95% confidence interval of the rates is ±10–13 points (Table 4.7); the differences between 57% and 70% should be read within this uncertainty. The run was carried out with a single seed; Feinberg and Rubright (2016) recommend that the number of replications be chosen according to the desired precision and justified.

The ranking steps of form v5 have not been tested. Because the run was carried out with form v4, the scoring of the `self_place` and `rank` steps (`SelfPlacement`, `RankScorer`) and the three-way comparison that places the self-view, the expression profile and the measurement side by side in the report lie outside the scope of Gate 0. These steps will be monitored under a separate heading in the Gate 1 report, and the desirability balance of the 64 statements will be measured in the pilot (ADR-044 Rationale and Impact).

The simulation does not confer validity. The last and most important limit is one of principle: because the persons were generated from Rehnüma's own model, recovery says nothing about what the measurement measures *in real Turkish-speaking people* (ADR-029 Rationale). The lasting value of the simulation is that it is a baseline: if label stability turns out low in the Turkish pilot, there is a point of comparison at hand for separating how much of the instability is "the translation is broken" and how much is "the label is fragile anyway" (validation protocol report, "Sonuç" [Conclusion]).

4.7 Analysis plan for Gates 1–4

The analyses of Gates 1–4 are carried out with a single toolchain (ADR-029 §1). The test service exports the anonymous response matrix in JSON Lines format: `php artisan pusula:export-responses dosya.jsonl --completed`. Each line is one test and contains the item metadata (section, type, stage, code, facet, key, expected answer), the raw answers, the page times, the quality output and the summary scores; in place of an identity, the first 12 characters of the SHA-256 digest of the test token are written (`Pusula-API/app/Console/Commands/ExportResponses.php`). A test checks that the export is anonymous and analysable (`test_disa_aktarma_anonim_ve_analiz_edilebilir`). The analysis is carried out with `docs/research/scripts/psychometrics.py`; the script uses only numpy and deliberately contains no psychometrics package, "so that what is being computed stays explicit". Table 4.9 maps the analyses of the script to the relevant gate and the decision criterion.

Table 4.9. `psychometrics.py` analyses, the relevant gate and the decision criterion.

Analysis	Function	Method	Gate	Criterion
Item matrix	`item_matrix`	Reverse-keyed items 6 – answer; instructed-response items excluded; with `--stage 1` only stage 1 items	1–3	—
Internal consistency	`cronbach_alpha`	Listwise complete rows; for facet, domain and RIASEC scale	1, 2	Facet α ≥ .60, domain α ≥ .80 (ADR-029)
Omega	`omega_total`	Single-factor principal axis; initial communality SMC, at most 25 iterations; for domain and RIASEC scales (not computed for facets)	1, 2	Reported together with α
Two-item facet projection	`spearman_brown`	`alpha_if_4_items`: projection from the stage 1 alpha to four items	1	Inter-item r ≥ .30 can proceed (note)

Analysis	Function	Method	Gate	Criterion
Item statistics	`item_stats`	Corrected item–total r, mean, SD; mean inter-item r	1	r ≥ .20 (ADR); the script lists < .30 as weak; negative r = keying error
Big Five structure	`big_five_recovery`, `efa`, `tucker_congruence`	30 facet means; principal components + varimax, 5 factors; matching to the target matrix by 5! permutations; Tucker φ per factor; facets loading on the wrong domain	2	φ ≥ .85 (ADR); at least 50, preferably 300 tests
RIASEC circular order	`riasec_order_test`	72 order predictions from the hexagonal distance of the 15 pairs of areas; proportion met; p from 2,000 label permutations	2	p < .05 (ADR); at least 30 tests
Quality	`analyze` (quality section)	Flag rates and distribution of levels	1	Flag rate < 20% (report)
Acquiescence	`acquiescence_report`	Mean, SD, rate of |index| ≥ 1.0 (all and under 18, at least 10 persons), correlation of the index with the five domains	1	Under 18, rate > 10% or |r| ≥ .30 → `reopen_adr_030`

Note. Source: `docs/research/scripts/psychometrics.py`. The criterion column is taken from ADR-029 and ADR-030; the entries marked "(note)" and "(report)" are taken from the research documents.

4.7.1 Gate 1: pilot item analysis

Gate 1 is carried out with about 100 real users and seeks answers to four questions at item level. The first is keying error: a negative item–total correlation after reverse coding shows that the translation has changed direction; because Gate 0 has ruled out keying errors at code level, a negative correlation seen in the pilot points directly to the translation. The first candidates are the items for which a shift of meaning was identified in 3.4.2. The second is facet and domain reliability. For two-item facets the inter-item correlation is reported instead of alpha; by Spearman–Brown, for the four-item alpha to be .70 the two-item alpha must be at least .54, that is, the inter-item correlation at least .37, and the lower bound that "can proceed" is r ≥ .30 (`dogrulama_protokolu.md` §2). For four-item facets omega should be given alongside alpha (Dunn et al., 2014); the script currently computes omega only for the domain and RIASEC scales, and at facet level gives alpha, the projection to four items and the mean inter-item correlation. The Turkish alphas should be read side by side with Johnson's (2014) values so that translation error can be separated from the weakness of the source: O6, O3, C3, A2 and A4 are expected to turn out weak in Turkish as well. The third is the rate of careless responding and the recalibration of the thresholds on Turkish and mobile data. The fourth is the decision point of ADR-030: the distribution of the acquiescence index, the flag rate under the age of 18 and its correlation with the domain scores; the script produces a `reopen_adr_030` field directly for this decision. Because the two ranking steps added by form v5 (`self_place`, `rank`) are a measurement separate from the Likert bank, they do not disturb the pilot comparison, but they are monitored under a separate heading in the Gate 1 report; the desirability balance of the 64 statements is also measured in this pilot (ADR-044 Rationale and Impact). The analysis of these steps is not yet in `psychometrics.py`.

The pilot also carries two monitoring obligations. Because the machine and craft items of the Realistic area (R:02, R:05, R:09) are perceived as gender-laden in Turkey, item means by gender should be reported (`ceviri_ve_madde_kalitesi.md` §4). The effect of the review screen should also be measured; however, the service overwrites section answers when they are resubmitted (`AssessmentController::saveSection`) and the export contains only the current answers. Because the

answers given before review are not stored, this effect cannot be measured with today's data (inference from the code); the research report's recommendation that "both the old and the new response are stored" should be implemented before the pilot.

4.7.2 Gate 2: structure

Gate 2 is carried out with at least 300 tests. For the Big Five, strict confirmatory factor analysis is not imposed: it is known that on the NEO-FFI ESEM fits markedly better than CFA and that CFA's assumption of zero cross-loadings is too restrictive for personality data (Marsh et al., 2010); the Chinese and Japanese versions of the NEO PI-R gave poor fit in CFA while clearly replicating the five factors with target rotation (McCrae et al., 1996). The proposed test is a five-factor solution on the 30 facets, target rotation according to Johnson's facet–domain assignment, and the Tucker congruence coefficient per factor; .85–.94 is counted as "fair" and ≥ .95 as "equal" similarity (Lorenzo-Seva & ten Berge, 2006). For RIASEC, a circular order test is carried out on the correlation matrix of the six scales; the CI = .69 that O*NET found for the IP-SF is the reference, and because in the cross-cultural literature the circular order is mostly supported while the strict equally spaced hexagon rarely is, the failure criterion should be the order.

Note. The script's implementation departs from this proposal at two points. First, the `efa` function uses principal components + varimax rotation and matches the factors to the target matrix by the best permutation; it is not the proposed Procrustes target rotation. Because the congruence of the varimax solution with the target will be lower than or equal to that of the Procrustes solution, the Tucker coefficients given by the script can be read as a conservative lower bound (inference). Second, the `correspondence_index` returned by the `riasec_order_test` function is the *proportion* of predictions met (0–1). In the literature the Hubert–Arabie correspondence index is usually defined as (met − violated) / total and ranges from −1 to +1; if this definition is correct, O*NET's .69 corresponds to about .85 on the script's scale and the "CI ≥ .60" criterion cannot be applied directly to the script output (inference; the definition should be verified from the primary source, Tracey 1997). Because the Gate 2 criterion of ADR-029 is only $p < .05$, the decision is not affected by this difference, but the scale of the number to be reported should be stated.

4.7.3 Gate 3: stability and external validity

Gate 3 is conducted through an optional retest invitation to users who have completed stage 2. ADR-029 prescribes at least 60 persons and an interval of 4–8 weeks; the criteria are a domain retest r ≥ .80 and a rate of the motif remaining the same of ≥ 60% (≥ 85% excluding balanced letters). Label stability should be reported as a percentage plus Cohen's kappa, both in the whole sample and in the "clear" subgroup (`dogrulama_protokolu.md` §4). The Gate 0 baseline is the point of comparison for this analysis: in the real retest, motif stability is expected to fall below 57%; how far it falls gives information about the size of real state variability.

For external validity, three types of evidence that a small team can collect have been planned. For convergent validity there are freely available benchmark scales in Turkish: the Turkish adaptation of the Ten-Item Personality Inventory (Atak, 2013) and the Turkish adaptation of the BFI-2 (Cemalcılar et al., 2021); the target is r ≥ .60 for the same domain (.50 with the TIPI) and ≤ .30 for a different domain. For self–other agreement the band of .31–.58 reported by Johnson (2014) is the reference; the target is r ≥ .40 per domain. For RIASEC, d ≥ .5 is expected for known groups (engineering → Realistic/Investigative, art → Artistic, teaching and health → Social), and the gender differences are expected to be in the same direction as in O*NET. That the correlation between occupation recommendation and user liking turns out small (r ≈ .10–.20) is consistent with the literature and should not be counted as a failure (Hoeglund & Hansen, 1999). These analyses are not yet in the script.

4.7.4 Gate 4: norms and fairness

Gate 4 opens Turkish percentiles banded by age and gender with at least 1000 tests. In classical norming EFPA counts N < 200 as inadequate, 200–299 as adequate, 300–999 as good and ≥ 1000 as excellent; in continuous norming, with eight subgroups, 70–99 persons per group is adequate, 100–149 good and ≥ 150 excellent, and in practice cNORM, with about 100 persons per cohort, gives an accuracy equivalent to 400 persons in classical norming (EFPA, 2013; `dogrulama_protokolu.md` §5). Comparison across groups can be made only after measurement invariance has been shown (ITC, 2017; SSI-1/2), and DIF analyses require at least 200 persons per group. Because of the large gender differences in the Realistic and Social areas in RIASEC, the decision between combined and separate norms should be written in an ADR: a combined norm reproduces the gender pattern in occupation recommendations, while a separate norm changes the meaning of the comparison.

Gate 4 is still open today; however, the source of the percentile in the report has changed independently of this gate. With ADR-082 (22 November 2026), personality percentiles come not from the reference group made up of those who took the test on Rehnüma but from the published open data of the scale: Johnson's (2014) IPIP-NEO-120 sample, 619,150 persons. The tables were recomputed from the same data separately for stage 1 (60 core items) and stage 2 (120 items); the mid-rank percentile of every possible total score is stored, and a normal distribution is not assumed. Of the 120 items, 117 matched the original key; because the three rewritten O6 items are not present in the data, no percentile is given for the O6 facet and the O domain is normed with the remaining original items (11 of 12 items at stage 1, 21 of 24 at stage 2). No percentile is shown for the interest areas; only the 0–100 score is given (`Pusula-API/app/Profile/PublishedNorms.php`; ADR-082 §1–5). This is not the Turkish norm that Gate 4 calls for, and it does not close the gate. The limits recorded by ADR-082 are these: the norm comes from volunteers who completed the English form online, and carrying it over to the Turkish adaptation is an assumption; the sample contains all ages and countries together, and separate norms by age and gender have not been used; the international means are above 50 (53–67 in the domains), so the same score may fall at a lower percentile than against the old reference group. A Turkish norm banded by age and gender, measurement invariance and DIF analyses remain the work of Gate 4.

The second criterion of Gate 4 is a pruning of content: in the "Does this fit you?" feedback in each section of the report, a sentence whose "doesn't fit" rate exceeds 40% is removed (ADR-029 §1; ADR-028 §5). This criterion is not sufficient on its own, because Barnum sentences are found to "fit" by everyone; for discriminative power one must look at whether the sentence is also found to "fit" by persons with the opposite profile (ADR-028 §5). This analysis rests on the `report_feedback` data on the Rehnüma side and is taken up in Chapter 5.

4.8 Tying the language of reporting to the gates

The basic rule of ADR-029 is that the claim of a gate that has not been passed does not enter the report. This rule determines the product's present language: because only Gate 0 has been passed, the report speaks within the limits permitted by the claim "the scoring is correct". Table 4.10 maps the language opened by each gate to its present implementation.

Table 4.10. Report language permitted by gate and the present implementation.

Gate	Claim opened	Permitted presentation	Present implementation
0 (passed)	The scoring code is correct	Raw 0–100 score; within-person ranking; "first impression" for stage 1; balanced letter and alternative motif; "neck-and-neck"	The `provisional`, `certain`, `alternatives`, `top_two_close` fields
1 (pending)	Pilot version, clean at item level	Facet bands in "tendency" wording	Facet bands are already in "tendency" wording; no pilot result
2 (pending)	The structure has been replicated in Turkish	Provisional percentile; removal of the "preliminary" label from the archetype and the code	Personality percentile, independently of the gate, from the published international sample (Johnson, 2014; 619,150 persons; ADR-082); no percentile for interest areas
3 (pending)	Stability and external validity	Individual report with confidence intervals (±1 SEM); T score	None
4 (pending)	Norms and fairness	"Turkish reference norm"; strengthened occupation recommendation	No claim of a Turkish norm; the source of the percentile is named as the "published international sample" (ADR-082)

Note. The "Permitted presentation" column is compiled from the gate table of the validation protocol report (the "Claim opened" column) and from ADR-029.

Note. There is a tension in row 2 of the table. The research report ties the provisional percentile to Gate 2 ($N \geq 300$); the product, however, shows the percentile without waiting for any gate. In version 1.1 this was done with the reference group of ADR-032: the percentile was given, in a group of at least 30 persons, in the form "higher than P% of the N people who took it on Rehnüma". ADR-082 changed the source: because the group is small, not randomly selected and contains trial sittings, personality percentiles are now computed from Johnson's (2014) published international sample, and the report names the source in an affirmative sentence: "personality percentiles are relative to the published international sample of the IPIP-NEO-120 (Johnson, 2014; 619,150 people)". `ReferenceGroup` remains in the code but is not connected to the report (ADR-082 Impact; 3.10). The tension thus changes form: the percentile no longer carries the sampling error of a small group, but neither does it mean that the Turkish structure has been replicated (Gate 2) or that a Turkish norm has been established (Gate 4); carrying the norm of the English form over to the Turkish adaptation is an assumption (4.7.4).

The language of evidence is itself tied to the gates. The expressions "validated test" and "scientifically proven" cannot be used, because the Turkish version has not passed Gates 1–4; the O*NET licence likewise excludes presenting the product as "validated" before the validation study is completed, as well as the expression "O*NET approved" (research report, "İlgi tarafında" [On the interest side]). The sentence that is defensible today is this: Rehnüma character analysis is built on published, open-source instruments tested in large samples; the scoring code has been verified by simulation; the psychometric validation of the Turkish version is continuing with the pilot, and Turkish reliability figures will be published in the technical appendix as they are collected. The validation protocol report notes that this honesty is a differentiator in its own right, since there is no B2C competitor in Turkey that publishes reliability figures.

The presentation of the report within the product (the distinction between the summary and the full report, the place of the trust note, the visual language of uncertainty) is the subject of Chapter 5. The principle that this chapter hands over to it is this: every degree of certainty in the report must be traceable by the name of the gate that stands behind it.

Chapter references

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/open-access-files.html
- Atak, H. (2013). On Maddeli Kişilik Ölçeği'nin Türk kültürüne uyarlanması. *Nöropsikiyatri Arşivi, 50*(4), 312–319. https://doi.org/10.4274/npa.y6128
- Cemalcılar, Z., Baruh, L., Kezer, M., Soto, C. J., Sümer, N., & John, O. P. (2021). Testing the BFI-2 in a non-WEIRD community sample. *Personality and Individual Differences, 182*, 111087. https://doi.org/10.1016/j.paid.2021.111087
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412. https://doi.org/10.1111/bjop.12046
- EFPA. (2013). *EFPA review model for the description and evaluation of psychological and educational tests* (Version 4.2.6). https://cms.bps.org.uk/sites/default/files/2022-07/110c_-_efpa_boa_test_review_model_version_4_2_6_approved.pdf
- Evers, A., et al. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema, 25*(3), 283–291. https://doi.org/10.7334/psicothema2013.97
- Feinberg, R. A., & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice, 35*(2), 36–49. https://ncme.org/wp-content/uploads/2025/10/Module-42-Simulation-Studies-in-Psychometrics-Fei.pdf
- Glavin, K. W., & Savickas, M. L. (2011). Interpreting Self-Directed Search profiles: Validity of the "Rule of Eight". *Journal of Vocational Behavior, 79*(2), 414–418. https://doi.org/10.1016/j.jvb.2011.01.005
- Gregory, C., & Lewis, P. (2016). *Linking client assessment profiles to O*NET occupational profiles within the O*NET Interest Profiler Short Form and Mini Interest Profiler (Mini-IP)*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/Mini-IP_Linking.pdf
- Hoeglund, T. J., & Hansen, J.-I. C. (1999). Holland-style measures of congruence: Are complex indices more effective predictors of satisfaction? *Journal of Vocational Behavior, 54*(3), 471–482. https://doi.org/10.1006/jvbe.1998.1675
- International Test Commission. (2017). The ITC guidelines for translating and adapting tests (2nd ed.). *International Journal of Testing, 18*(2), 101–134. https://doi.org/10.1080/15305058.2017.1398166
- Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*(2), 202–220. https://doi.org/10.1177/1094428105284919
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57–64. https://econtent.hogrefe.com/doi/10.1027/1614-2241.2.2.57

- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19–40. https://psychology.psy.sunysb.edu/attachment/measures/content/maccallum_on_dichotomizing.pdf
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*(3), 471–491. https://doi.org/10.1037/a0019227
- McCloy, R., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999b). *Development of the O*NET computerized Work Importance Profiler*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/DevCWIP.pdf
- McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 70*(3), 552–566. https://doi.org/10.1037/0022-3514.70.3.552
- National Center for O*NET Development. (n.d.-d). *O*NET Tools Developer License*. https://www.onetcenter.org/license_toolsdev.html
- Pittenger, D. J. (1993). The utility of the Myers-Briggs Type Indicator. *Review of Educational Research, 63*(4), 467–488. https://doi.org/10.3102/00346543063004467
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of Big Five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718–737. https://doi.org/10.1037/0022-3514.94.4.718
- Tracey, T. J. G. (1997). RANDALL: A Microsoft FORTRAN program for a randomization test of hypothesized order relations. *Educational and Psychological Measurement, 57*(1), 164–168. https://doi.org/10.1177/0013164497057001012
- Welz, M., & Alfons, A. (2023). *When respondents don't care anymore: Identifying the onset of careless responding* [Preprint]. arXiv. https://arxiv.org/abs/2303.07167

5. Reporting the Results

The validity of a measurement instrument is not limited to the accuracy of its scores; the interpretations derived from those scores must also be defensible. The standards of AERA, APA and NCME require score information to be presented in language suited to the target audience, stating what the test covers, what the scores represent and how precise they are (Standard 6.10; the full text of the standards could not be accessed, secondary citation). Because Rehnüma Character Analysis produces a fully automated report with no counsellor and no supervision, this second validity problem weighs more heavily here. In the classification of the International Test Commission (ITC, 2005), the instrument operates in "open mode": the person takes the test out of personal interest, their identity cannot be verified, and there is no expert to whom they can put a question about the result. The report has to carry, on its own, the interpretation an expert would provide.

This chapter describes how the report is built. We first set out the reporting principles and the literature on which they rest (5.1). We then cover the opening of the result behind an account and the distinction between the summary and the full report (5.2), the information architecture of the report (5.3), the norm against which percentiles are given and the language used for them (5.4), the use of universal codes together with their Turkish equivalents (5.5), the visualisation decisions (5.6), the trust design (5.7) and the pruning of report texts through user feedback (5.8). The measurement design (Chapter 3) and the validation protocol (Chapter 4) are taken as given here; the character trace, that is, the transfer of the result to the dashboard, is the subject of Chapter 6. Every claim about the code is tied to a file in the repository and every decision to the relevant ADR; the state described is the code as of 22 November 2026, and the places where it departs from Version 1.1 are noted as history. Design rationales not supported by the literature are marked as "inference".

5.1 Reporting principles

5.1.1 The Barnum effect and the limits of perceived accuracy

One of the oldest and most robust findings in the personality feedback literature is that people accept, with high accuracy ratings, statements that are general, positive and presented as "specific to you". In Forer's original study, 39 students rated a general personality description, given identically to everyone, at a mean accuracy of 4.3 on a 0–5 scale (Forer, 1949; secondary citation). Most of the subsequent literature relies on Forer's set of 13 statements. Dickson and Kelly's (1985) review shows that the acceptance of such profiles is determined mainly by two properties: that the profile appears relevant to the person, and that it is positive. The generality of the interpretation, its source, its format and the characteristics of the administrator have also been examined, but were not found to be as decisive as the first two. High base-rate interpretations, that is, those true of most people, receive higher accuracy ratings than low base-rate ones.

The consequence of this finding for product design is serious. In a NEO-based feedback study, Poškus et al. (2014) had participants rate their real summary, a general summary and a socially desirable summary; they found no difference in perceived accuracy among the three and showed only that the

reversed summary was judged less accurate. In a small, non-peer-reviewed poster (n = 89), those who received a general result found the test more accurate than those who received a specific result (Oklahoma Baptist University, 2023; evidence grade weak). Changing the visual format of IPIP feedback has been reported not to predict perceived accuracy, whereas perceived face validity did predict it (Studia Doctoralia, reference to be verified). In a study of computer-based test interpretation (CBTI), psychologists who received a computer report made poorer judgements on an ambiguous MMPI-2 profile than those who did not (secondary citation). Read together, these studies say the following: **perceived accuracy is not evidence of the validity of the report.** A high rating in response to "How well did this report describe you?" may be a metric that rewards the Barnum effect. For this reason, perceived accuracy is not used as a quality criterion in Rehnüma; feedback is collected only to prune sentences, as a one-way filter (5.8).

With respect to behavioural consequences the picture is mixed: in one doctoral thesis the Barnum effect occurred but self-reported behavioural tendencies did not change (University of Warwick, n.d.), while the effect of false feedback on mood was found to depend on the credibility of the test (University of Oxford, n.d.). It follows that the report text can affect how a person views themselves, and that a general, positive text is not a harmless ornament (inference).

5.1.2 Four writing rules

We gathered the countermeasures that emerge from the literature into four rules. The rules are written in the note at the head of `Pusula-API/database/data/report_texts.json`, the single source of the report texts in Pusula-API, and are part of ADR-026 and ADR-028.

Rule 1: Every sentence is tied to a score. ITC (2005) guideline 25.4 requires that the rationale for automated interpretive statements, and the score or score pattern from which they are derived, be explained. The EFPA test review model likewise requires that the "granularity" of the report, that is, the number of distinct bands on a scale that are mapped to text, be justifiable by the measurement error of the scale (EFPA, 2025, §12.1). In Rehnüma the report text is tied to scores at three levels: the archetype text to the domain profile, the strength and growth cards to facet bands, and the tendencies to facet and RIASEC formulas. At the facet level only three bands are used (low, medium, high; ADR-026 §6). Sentences that describe the person are not generated in the front end; as the note at the head of `ReportView.tsx` says, they come from Pusula and are only laid out here. The front end holds only fixed framing sentences such as the quality note and the line "Kendini … olarak gördün; ölçüm … diyor" (You saw yourself as …; the measurement says …). The report text is stored keyed in Pusula and is dressed in the language of the request at read time (Turkish, English, French, German, Arabic; ADR-046, ADR-058); the scores do not change with the language.

Rule 2: Both poles are written as strengths. Writing the two ends of a dimension as "good" and "bad" is both an ethical problem and something that enlarges the Barnum risk: a text that describes only the positive pole becomes acceptable to a low scorer as well. Each facet record therefore carries the same four fields for both poles: `strength`, `cost`, `tip` (a suggestion to try) and `looks_like` (how it appears in everyday life). For example, the high pole of facet S1 (Managing worry) is named "Rahat" (At ease); its strength is written as "olası sorunlara takılıp kalmamak" (not getting stuck on possible problems) and its cost as "riskleri hafife alıp hazırlığı son ana bırakmak" (underestimating risks and leaving preparation to the last minute); the low pole is named "Kaygılı" (Anxious) and its strength is "olası sorunları önceden görüp tedbir almak" (seeing possible problems in advance and taking precau-

tions). In the front end, the pole names of the five domains were chosen on the same principle (`Front-end/messages/tr/report.json`, `traitPoles`): for Openness to experience "Alışılmışı seven" (Likes the familiar) and "Yeniyi arayan" (Seeks the new), for Conscientiousness "Akışına bırakan" (Goes with the flow) and "Planlı" (Planned), for Emotional stability "Hassas" (Sensitive) and "Dingin" (Calm). No pole name implies a deficiency.

Rule 3: The cost of every strength and the advantage of every growth area are written. Dickson and Kelly's (1985) positivity finding points directly to this. Purely positive language increases acceptance and lowers discrimination. When cost and benefit are written together, the sentence becomes both more specific and falsifiable: the person can say "I do not have this cost".

Rule 4: The language is situational and behavioural. Instead of adjective sentences such as "You are creative", the likely behaviour in a specific situation is described. The `looks_like` field exists for this purpose: the sentence "Bir sınavdan ya da toplantıdan önce olası senaryoları kafanda defalarca kurarsın" (Before an exam or a meeting you run through the possible scenarios in your head many times) was written for low S1 and is plainly false for someone high on S1. This is the workable form of the base-rate test.

The check corresponding to each rule is given in Table 5.1.

Table 5.1. The Barnum checklist and its implementation in Rehnüma.

Check	Basis	Counterpart in the implementation	Status
Base-rate test: a sentence is deleted if it is also true for a person with the opposite score	Dickson and Kelly (1985); high acceptance of high base-rate interpretations	Note in `report_texts.json`: no sentence was written that would also be true for a person with the opposite score	Manual at the writing stage; empirical test at Gate 4
Both poles together, both as strengths	ITC (2005) 25.4; scoring and reporting note	`high` and `low` records for every facet; `traitPoles`	Implemented
Cost–benefit pair	Positivity finding (Dickson & Kelly, 1985)	`strength` + `cost` fields; "bedeli" (its cost) on the strength cards	Implemented
Situational, behavioural language	Scoring and reporting note, S4 inferences	`looks_like` field	Implemented
Granularity matched to measurement error	EFPA (2025) §12.1–12.2	Only 3 bands for facets; a distance of at least 15 points from the middle for a personal line	Implemented; threshold to be re-established with data
Number of strengths varies with the profile	The template impression of a fixed 9/9 layout (inference)	Strengths drawn from facets clearly away from the middle; number variable	Implemented
Upper limit on length: about one page per scale	EFPA (2025) §12.6	Collapsible sections; detail closed by default	Partial; see 5.3.3
Statement-level accuracy feedback	Poškus et al. (2014)	"Bu sana uyuyor mu?" (Does this fit you?) buttons, `report_feedback`	Being collected; analysis at Gate 4

Note. The "Status" column is written against the code in the repository as of 22 November 2026. The report texts themselves have not yet been validated (`Pusula-API/AGENTS.md`).

ADR-028 adds a structural layer to these rules: **the label is the door, the content is the person**. The report speaks through 16 mind motifs and 4 temperaments, but every section consists of two layers. The first layer is the base text of the motif or temperament. The second layer consists of the "In your answers" lines that come from the person's facets relevant to that section. A personal line is written

only if the facet shows a clear departure from the middle, that is, a distance of at least 15 points (`Pusula-API/app/Profile/ProfileBuilder.php`, `DISTINCT = 15.0`); otherwise the report says "you are balanced in this area". The rationale is plain: if two people carrying the same motif label read only the base text, they would have read the same text, and this would be the Barnum risk in institutionalised form. The facet lines distinguish two people with the same label from each other. They also prevent the report from changing wholesale if the label changes at retest; the facet lines remain, independent of the label. The life-area sections added by ADR-045 do not carry this second layer (5.3.2).

5.1.3 Communicating uncertainty

Every score carries measurement error. In classical test theory the standard error of measurement is given by SEM = SD·√(1 − reliability); with a reliability of .80 the SEM is roughly 0.45 SD, and a band of ±1 SEM can correspond on the percentile scale to an interval as wide as from the 25th to the 60th percentile (scoring and norms note, S1 inferences). EFPA (2025, §12.2) requires that the interpretation of scores and score differences be defensible in the light of measurement error: a small difference between two dimensions may not be meaningful. Hiding uncertainty makes the report look more precise but makes it less honest.

The Gate 0 simulation run with form v4 (ADR-029; Chapter 4) showed numerically in which layers the uncertainty accumulates. When synthetic persons whose true traits never changed retook the test, the four-letter code stayed the same only 57% of the time. Clear letters never flipped; 36% of "balanced" letters flipped. The first letter of the Holland code stayed the same 45% of the time, but one of the first two letters was retained 75% of the time. The first archetype stayed the same 70% of the time. These figures are an unavoidable consequence of cutting continuous scores into categories (McCrae & Costa, 1989): a score close to the boundary crosses to the other side with a small measurement error. The solution adopted in ADR-029 is not to hide the code but to carry the uncertainty. The corresponding signals in the report are gathered in Table 5.2.

Table 5.2. Sources of uncertainty and their signals in the report.

Source	Criterion	Signal in the report	Code
A trait-code letter is close to the boundary	Distance of the domain score from 50 < 8 → "Balanced"; < 20 → "Moderate"; otherwise → "Clear"	A clarity label on every axis; if there is a balanced letter, the neighbouring motifs are listed as "alternatives"	`ProfileBuilder.php` (`BALANCED = 8.0`), `pages/Motif.tsx`
The top two Holland areas are neck-and-neck	Difference between the top two areas ≤ 5 points → `top_two_close`	"Your top two areas are neck and neck; the first two letters of your code may swap places on another day"	`ExpressedInterest`, `pages/Interests.tsx`
Facets were measured with 2 items (stage 1)	`provisional = stage < 2`	"In your answers · first impression"; in the technical appendix "when you go deeper it rises to 4 and becomes firm"	`parts.tsx`, `TechnicalNote.tsx`
Archetype assignment	Similarity percentages to 8 types, summing to 100	A list with percentages instead of a single label; if the first two types are within 10 points, "read both"	`pages/Archetypes.tsx`, `pages/Overview.tsx`

Source	Criterion	Signal in the report	Code
Response quality	Four indicators + the acquiescence index; levels `ok`, `note`, `review`, `uninterpretable`	A confidence note by level on the At a glance page; in the technical appendix a readable quality label and a plain-language explanation of the flags (ADR-044 §4)	`pages/Overview.tsx`, `TechnicalNote.tsx`; texts in `report.json` (`overview.quality`, `technical.flags`)
A scale with no counterpart in the published norm	Facet O6; the six interest areas	No percentile; only the 0–100 score (ADR-082 §4–5)	`PublishedNorms.php`, `pages/Interests.tsx`
Measurement error (SEM)	No Turkish reliability yet	Not shown; reliability ranges of the original instruments in the technical appendix	Open task (after Gate 3)

Note. The clarity thresholds and the balanced threshold were chosen according to the retest difference in the simulation (SD ≈ 6–7) (ADR-029 §2); they will be re-established with real user data.

The quality level is communicated in three-step language. At the `note` level the report says "There is a small inconsistency in your answers; the results still describe you to a large extent". At the `review` level the warning "read the results with caution" appears. At the `uninterpretable` level the report says that it may describe the person incompletely and suggests retaking the test at a calm moment. This language is the reporting counterpart of the "no blocking" decision in ADR-026 §5. The literature shows that making a hard decision on a single indicator would also penalise attentive respondents: among 12,578 upper-secondary students, a strict threshold flagged 37% of participants on at least one indicator, and most of those flagged were caught on a single indicator only (Chauliac et al., 2023). The flag texts in the technical appendix are not accusatory either; they are observation sentences such as "the same option was marked for a long stretch" (scoring and norms note, S7).

The last row of Table 5.2 records an open gap: a confidence interval for the person's own score is not yet shown, because the reliability coefficients of the Turkish version are not known. The validation protocol ties the ±1 SEM interval in the individual report to the period after Gate 3 (Chapter 4). Until then, uncertainty is carried by categorical signals (balanced, neck-and-neck, first impression); these signals do not take the place of a continuous interval (inference).

Explaining stability to the user is a separate matter of importance. For Big Five domain scores, six-month retest correlations in BFI-2 adaptations are at the level of r ≈ .86, and for facet scores r ≈ .80; in adolescence rank-order stability increases with age, and interests stabilise markedly between the ages of 18 and 22 (Low et al., 2005; Pullmann et al., 2006). The recommended language is the frame "this is a photograph, not a seal" (scoring and norms note, S6). In the report this principle appears in two places: the alternative motifs listed on the motif page for a balanced letter, and the item in the technical appendix that reads "The trait code, temperament and types are labels on continuous scores". The note beneath the axis cards, "If the dot is close to the middle, the letter may flip on another day", was removed from the page in the 22 November simplification (5.3.3). The invitation recommended for adolescents, "retake it in 1–2 years", is not yet in the report and remains an open task.

5.1.4 Language tied to the gates

The most important provision of ADR-029 from the standpoint of reporting is this: **a claim that belongs to a validation gate not yet passed does not go into the report.** The five gates (0 simulation, 1 pilot, 2 structure, 3 stability, 4 norm) are not only a research plan but the version control of the report's language. Table 5.3 shows which statement stands behind which gate.

Table 5.3. Validation gates and the statements they unlock in the report's language.

Gate	Status (22 November 2026)	Language used today	Language unlocked when the gate is passed
0 · Simulation	Passed	"Scoring has been tested"; facets are a "first impression" at stage 1	—
1 · Pilot (N ≈ 100)	Pending	"The Turkish adaptation is in its pilot stage"; reliability only for the original instruments	Turkish α and ω values in the technical appendix
2 · Structure (N ≥ 300)	Pending	Archetype and code are derived labels; prototypes set by hand	Archetype centres established by clustering; removal of the "preliminary" label
3 · Stability (N ≥ 60, 4–8 weeks)	Pending	"A letter near the boundary may flip on another day"	Retest rates in real users; ±1 SEM interval
4 · Norm (N ≥ 1000)	Pending	"Personality percentiles are relative to the published international sample of the IPIP-NEO-120" (ADR-082); no percentile for interest areas	Turkey-specific percentiles in age and gender bands; Barnum pruning

Note. The gate thresholds are taken from ADR-029 §1. The public trust page (`/analysis/trust`; text in `Front-end/messages/tr/trust.json`) shows Gates 1–4 as "Pending"; the "In progress" discrepancy recorded in Version 1.1 has been resolved.

The most visible consequence of this principle is at the level of wording. The word "doğrulama" (validation) is not used for the Turkish version (ADR-031 §2). Phrases such as "doğrulanmış test" (validated test) or "bilimsel olarak kanıtlanmış" (scientifically proven) appear on no surface of the product. In their place the formula "built on published, open-source instruments tested in large samples; the psychometric validation of the Turkish version is under way with the pilot" is used. The technical appendix says this with the names of the gates: "simulation passed; the pilot, structure, stability and norm gates are pending" (`Front-end/src/components/test/report/TechnicalNote.tsx`). The source of the percentile is also named: the trust page and the entrance page of the analysis say that personality percentiles are given relative to the scale's published international sample; a Turkey-specific percentile stands behind Gate 4 (ADR-082; 5.4).

Language tied to the gates offers one way of implementing the requirement of ITC (2005) guideline 29.1 and of APA Ethics Code 9.06 (American Psychological Association, 2002) that the limitations of interpretations be stated. What distinguishes it is that the limitation is stated not in a general warning box but right next to the claim concerned. The reader sees the "balanced" label next to the letter and the "neck-and-neck" note next to the interest code (inference: the assumption that a warning placed in context will be read more than a general warning has not been tested). The percentile is where this principle is not applied today: the source of the norm is not written next to the number (5.4.3).

5.2 Opening the result: account, summary and full report

5.2.1 The account step and the current flow

In Rehnüma the character analysis is the first value the product delivers; sign-up comes after it (ADR-025, Decision 1). The place of the account in the flow has changed three times. ADR-025 asked for the email before the test. ADR-026 §4 moved the email to the end: the test started anonymously, the

email was requested after the sections, a person who had not signed up saw an open summary, and the full report was opened with a password; Version 1.1 described this flow. ADR-073 (15 November 2026) removed the summary without sign-up: **the result is not shown without an account, scoring is done after the account is created, and the result appears on the dashboard.** The rationale: the result lived in two places, and because a person who had finished the test always landed on their result at `/analysis`, they could not start a new analysis.

The basis for requesting the account at the end of the test rather than at the beginning has not changed; it is the funnel literature, and its evidence grade lies between moderate and weak. No independent A/B experiment could be found; according to vendor data, a form placed immediately before the result obtains consent at a rate of 35–55% (vendor opinion). In a trial by the U.S. Census Bureau, break-off almost doubled on the screen requesting contact information (0.9% versus 0.5%; Zelenak, 2016). The person now creates an account without seeing any part of their result; the effect of this change on the break-off rate has not been measured (inference).

The current steps of the flow are as follows (`Front-end/src/components/test/TestFlow.tsx`):

1. **Entrance page** (`/analysis`; `entrance/Entrance.tsx`): instrument names, item counts, sources, limits and what happens to the data; the single action is "Start the analysis".
2. **Situation and consent** (`steps/About.tsx`): life stage (upper-secondary school, university, working, searching), year of birth, optional target occupation (chosen from the catalogue, free text is not accepted; ADR-074), acknowledgement of the privacy notice, parental consent under the age of 18 (Chapter 8.1).
3. **Stage 1 sections**: the self-placement step (eight portraits; ADR-044), 60 personality items, 30 interest items, the choice of expressed interest and one instructed item (ADR-026 §1).
4. **Optional review** (`steps/ReviewPage.tsx`): when the quality level comes out as `review` or `uninterpretable`, only one item of each contradictory pair is asked again, in shuffled order and without the previous answer being shown (ADR-026 addendum).
5. **Account** (`steps/Register.tsx`): email, password and optional name, or continuing with Google (ADR-065). A person who is already signed in does not see this step.
6. **Transition screen** (`steps/Building.tsx`): the bridge record is linked to the account, the test is scored and the result is copied to the profile; the screen does not close before scoring ends.
7. **Dashboard** (`/overview`): the character analysis card shows the result and opens the report (`/analysis/report`).
8. Optional **stage 2** (`/analysis/deepen`, about 25 minutes): statement ranking (8 rounds × 8 statements), completion of the facets to 4 items, work values and rhythm.

The sales model has also changed. ADR-041 had defined a one-off "full report + dashboard" package and ADR-061 three tiers. ADR-083 (22 November 2026) reduced these to two tiers: the analysis is free and gives the open sections of the report together with the opening sentence of the locked sections; all sections of the report open only with the annual dashboard (TL 1,500; TL 1,200 at launch); the full report is not sold separately, and accounts that had previously bought the report keep their entitlement. The rule that promotional copy should say what is given instead of using the word "free" (ADR-041 §1) remains in force. The ethical side of the commercial model is in Chapter 8.6.

The back end does not yet enforce ADR-073; the ADR records this as a separate task. `Back-end/app/Services/PersonalityTestService.php` continues to return a summary to a bridge record that has no sign-up but does have an email written to it; on a bridge record without an email, `requireEmail` returns 422. `Back-end/tests/Feature/InvariantsTest.php` protects this old contract: the ADR-026 §4 test checks that the summary is not given without an email, and the ADR-025 test checks that sign-up is not tied to the test and that the summary remains limited to three occupations. The interface no longer calls these endpoints for a person who has not signed up.

5.2.2 Where the restriction lives: the server

The distinction between the summary and the full report is made on the server, not in the front end. The note at the head of the class `Back-end/app/Services/Pusula/PersonalityReport.php` says so explicitly: the restriction is applied here, and Pusula's own paywall flag is not trusted. The class has two methods: `summary` produces the restricted view, `full` everything that Pusula provides. The difference in content is given in Table 5.4.

Table 5.4. Content of the summary view and of the full report.

Area	Summary (`summary`)	Full report (`full`)
Archetype	Only the closest type: name, tagline, similarity percentage, general text	All eight types with their percentages and all their sections
Five domains	Score, band and domain sentence	The same; facet scores are also in the data
Percentiles	`n` and the domains	Domains and facets (except O6)
Self-view (ADR-044 §7)	The portraits chosen and their overlap with the measurement	In addition, the expression profile and the three-way comparison data
Interests	Holland code, hexagon, six areas	The same
Locked sections (motif, temperament, self type, relationships, parenting, at work, strengths, tendencies)	Only the first sentence of each (`openings`; at most 200 characters)	All content
Five branches	None	Selected cells per branch (Chapter 7)
Occupations	The top 3, without the fit score and its components (`slug`, `title`, `holland_code`, `band`, `group`, `job_zone`); fit with the target occupation (ADR-081 addendum)	The whole stored list (at most 100), challenging occupations, the target occupation card (ADR-074)
Quality	Level, readable label and flags	The same

Note. Source: `PersonalityReport::summary`, `PersonalityReport::full` and `PersonalityTestService::report`, `profile`. The raw quality indicators and the item pairs are not sent to the client in either view (the `quality()` method).

Which view is returned to whom does not, today, fully coincide with the two-tier model. `PersonalityTestService::report` gives the summary to a bridge record without sign-up and the full report to a bridge record linked to an account. The report endpoint for a signed-in user (`GET /test/profile`), however, returns the `full` view to every account; no check tying the view to the purchase entitlement (`entitlements`) could be found in the code as of 22 November 2026. Because the

result is not shown without an account (ADR-073), the summary view is today rendered in the interface only on the development page `/dev/report-summary`. ADR-083 counts redirecting the lock box to the dashboard as an open task; the box still says "Open the full report". Below, the summary view is described as the product behaviour defined by ADR-081 and ADR-083.

The rationale for applying the restriction on the server is security: a field hidden on the client remains visible in the network response; paid content that can be read with the browser's developer tools is both a leak and something that turns the paywall into decoration. Version 1.1 had also argued that this separation was a structural barrier against the design pressure towards "showing a little to arouse curiosity". ADR-081 showed that this pressure materialised: the locked section now opens and the person reads one sentence from their own result (5.7.2). The boundary nevertheless remained on the server. Only the first sentence of each locked section goes out to the summary; the name, the code, the score and the rest of the text of the section do not go to the client. `InvariantsTest` makes this executable: the JSON of the summary must contain neither a second sentence nor a code such as "ISTP".

The quality information is filtered as well. The `quality()` method passes only the level, the readable label and its explanation, the flag names and the `reliable` field; for the review screen, `publicCheck` gives only the level, the flags and the text of the contradictory statements. The reason for hiding the raw indicators is to prevent the person from "gaming" the indicators (inference).

In the summary the occupation list arrives without the fit score and, under the occupations addendum to ADR-081 (22 November 2026), is not shown in readable form: the list is rendered blurred, with the lock box beneath it. The only occupational information left open is the fit with the target occupation the person chose at the start of the test. With respect to the traceability requirement of ITC 25.4, the summary is weaker than the full report (inference); Chapter 8.6 addresses this tension from the standpoint of the commercial model.

5.2.3 The second stage and the two resolutions of the full report

The report seen after the account is created is produced from stage 1 data. This distinction is stated explicitly in the interface: the eyebrow label of the overview page is "Full report · first stage" or "Full report · two stages" (`pages/Overview.tsx`). At stage 1 each facet is measured with 2 items; at Gate 0, facet recovery at this level came out at a median of r = .81 and a minimum of r = .78 (ADR-029 §2). Stage 2 completes the facets to 4 items and recovery rises to a median of r = .89. The language of the report changes accordingly: at stage 1 the personal lines are marked "first impression", and the technical appendix says of the facets that "when you go deeper it rises to 4 and becomes firm". When stage 2 is completed, `PersonalityTestService::stageComplete` writes the result to the bridge record and to the `personality_profiles` table.

That stage 2 is optional is an explicit provision of ADR-031 §4: the entry screen of the second stage has the route "Not now, back to my report". With form v5 the second stage grew to about 25 minutes, and the 12-minute claim for the first stage was retained (ADR-044 §2). The design thus offers two resolutions of the same measurement (Rehnüma personality test psychometrics report, Conclusion).

5.3 Information architecture of the report

5.3.1 Paged structure: fourteen sections

In its first version the report was a single long page. In that form the user could not see "where they were", and ADR-032 moved the report to a paged structure of ten sections. ADR-045 (6 November 2026) added three sections devoted to areas of life (In relationships, In parenting, At work); together with the Five branches section (ADR-037; Chapter 7) the report today has fourteen sections. ADR-081 (22 November 2026) and its addendum of the same day changed the page behaviour: the outer page does not scroll, the report fills the rest of the screen, and only the content of the section scrolls within its own frame. The first solution, which measured the sections and split them into screens, was removed the same day because it produced 50 screens on desktop and 92 on a phone. On a wide screen there is a section navigator on the left, on a narrow screen a section picker; the previous and next buttons sit beside the navigator, and the `#bölüm` fragment in the address bar, the arrow keys and sideways swiping on touch devices do the same job (`Front-end/src/components/test/report/ReportPager.tsx`). A fixed language strip sits at the bottom of the screen; changing the language rewrites the report in that language (ADR-058). The list of sections is given in Table 5.5.

Table 5.5. The fourteen sections of the report.

#	Section (`id`)	Component	Data it rests on	In the summary
01	At a glance (`ozet`)	`pages/Overview.tsx`	Archetype, code, temperament, Holland code, domain percentiles, quality	Open (without code or temperament)
02	Five core traits (`kisilik`)	`pages/Personality.tsx`	Domain scores, their percentiles and domain sentences	Open
03	Mind motif (`motif`)	`pages/Motif.tsx`	Trait code, axes, clarity, alternatives, motif sections	Locked
04	Temperament (`mizac`)	`pages/Temperament.tsx`	Temperament key, O, C, A scores	Locked
05	Self type (`tipler`)	`pages/Archetypes.tsx`	Similarity to the eight archetypes, type details	Locked
06	Interest areas (`ilgi`)	`pages/Interests.tsx`	RIASEC scores, `top_two_close`	Open
07	In relationships (`iliski`)	`pages/LifeAreas.tsx`	The `romance` and `friends` texts of the closest archetype	Locked
08	In parenting (`ebeveyn`)	`pages/LifeAreas.tsx`	The `parenting` text of the closest archetype	Locked
09	At work (`iste`)	`pages/LifeAreas.tsx`	`work_roles`: employee, teammate, manager	Locked
10	Strengths (`guclu`)	`pages/Strengths.tsx`	Facets clearly away from the middle	Locked
11	Tendencies and values (`egilim`)	`pages/Drives.tsx`	Natural tendencies, work values, rhythm	Locked
12	Five branches (`besdal`)	`pages/FiveDomains.tsx`	Cells of the five-branch character matrix (Chapter 7)	Locked; no opening sentence
13	Occupations (`meslek`)	`pages/Careers.tsx`	O*NET occupation fit, target occupation	Target occupation open; list blurred

#	Section (`id`)	Component	Data it rests on	In the summary
14	Closing (`kapanis`)	`pages/Closing.tsx`	Direction of growth, suggestions to try, technical appendix	Open

Note. Source: the `chapters` array in `ReportView.tsx`; the open and locked distinction follows ADR-081 §7 and the occupations addendum. If there is no interest data, section 06 drops out of the list.

The paged structure has three functions: orientation (the sequence number in the navigator and the `aria-current` marker tell the person where they are), division of the information load, and transparency. In the summary, locked sections appear in the navigator with a lock icon and, for screen readers, with the suffix "(in the full report)"; the person sees page by page what they are missing (ADR-032 §1). How a locked section is displayed is discussed in 5.7.2.

The navigator is a `nav` element, the section buttons are real `button` elements, and the arrow keys do not operate while an input field has focus. Thanks to the fragment in the address bar, a direct link to a section can be given.

5.3.2 The two-layer section: base text and personal line

The inside of the sections is built with the two-layer structure of ADR-028. The motif and temperament sections consist of collapsible elements (`Disclosure`). Each collapsible element shows first the base text and then the "In your answers" lines that come from the person's relevant facets (`report/parts.tsx`, `SectionBody` and `PersonalLines`). The mapping between section and facet is fixed in ADR-028 §2: "Overall pattern" uses the two most pronounced of the five domains, the "At work" section the facets C1, C2, C4, C5, C6, E3 and O5, "Typical pursuits" O1, O2, O4, E4 and E5, "Relationships" A1, A3, A4, A5, A6, E1, E2 and S4, and "Under pressure" S1, S2, S3, S5 and S6; the "Suggestions" section is made up of the suggestions to try of the three most pronounced facets.

This mapping has two consequences. First, the content of every section is tied directly to the measured construct. EFPA (2025, §12.3) requires an assessment of how far the report goes beyond the information provided by the scores; a section such as "Relationships", which is only indirectly connected to the measured construct, becomes defensible only when it is tied to the relevant facets. Second, the absence of a personal line is itself information: the sentence "you are balanced in this area" says that there is no clear departure in the facets of that section.

The three life-area sections of ADR-045 meet this criterion less well. The texts are written per archetype (`report_texts.json`: `romance`, `friends`, `parenting`, `work_roles`; 70–100 words each), and the page shows the text of the closest archetype; in the relationships and parenting sections, if the second type is within 10 points, its text is added as well. The writing rhythm is fixed: an opening with the strength, a single point to watch introduced by "yine de" (still), and a balanced close (ADR-045 §2). There is no facet line in these sections; two people closest to the same archetype read the same text. The "the label is the door, the content is the person" principle of ADR-028 is applied here only at the archetype level, and the Barnum risk is higher than in the other sections (inference); testing the texts with feedback in the pilot is an open task (ADR-045, Impact).

The "under pressure" and "direction of growth" sections of the archetypes are the Big Five-based counterpart of the "self trajectory" concept that Corpitall took from the Enneagram (ADR-028 §4). These sections rest on the Emotional stability and Conscientiousness facets. The Enneagram itself is not

used, in accordance with ADR-026: the nine types were not derived statistically and the RHETI is a copyrighted instrument.

5.3.3 The length limit, collapsible elements and the 22 November simplification

EFPA (2025, §12.6) states that reports longer than an average of one page per scale may be over-interpreted. The full report of Rehnüma strains this limit in terms of volume of content: there are five sections for each of 16 motifs, six sections for each of four temperaments, the type details and life-area texts of eight archetypes, and three sub-fields on every strength card. The collapsible structure does not load this volume directly onto the reader; in the motif sections only the first element arrives open, the others are closed. But the collapsible structure does not solve the length problem: even if a sentence sits inside a closed element, it is a claim at the moment it is read and has to pass the Barnum check (inference). The feedback loop in 5.8 is the way of paying the price of this volume.

On 22 November 2026 the report was simplified by decision of the user (commits fc27cd4, fdc0ed1, fb3b811, 4f70ff2, 286a3bb). The method and "how to read this" explanations at the head of the sections, the breakdown of 30 facets on the Personality page, the table of four temperaments on the temperament page, the "what you said and what was measured" comparisons on the interest and At a glance pages, the tendencies in the middle of the ranking (the top three and the bottom two remain), the source references, evidence grade labels and method box in the Five branches section, and the disclaimer paragraph in the closing were removed. The three-way comparison that ADR-044 §3 had placed on the Self type page was reduced to a single sentence: if the portrait the person chose diverges from the measurement, this is said in one line; if it coincides, nothing is written. The simplification brings the report closer to EFPA's length criterion; its price is that facet scores remain only on the strength cards and in the personal lines, and that some notes on limitations have been separated from the claim they qualify (5.4.3, 5.5).

The collapsible element is a real button and carries the `aria-expanded` state (`parts.tsx`).

5.3.4 Technical appendix

On the closing page of every report there is a collapsible technical appendix (`report/TechnicalNote.tsx`; ADR-029 §1, ADR-031 §5). The "Measurement" heading states the instruments used, how many items each facet was measured with in this report, that the self-placement and statement ranking steps do not enter the score, and that the scores are on a linear 0–100 scale. The "Precision" heading states that the domain scores are the most stable layer, the reliability ranges of the original instruments (IPIP-NEO-120 domain α .81–.90, facet α .63–.88; O*NET Interest Profiler α .78–.87, retest .78–.86; work values retest .59–.66), that derived labels may flip for letters near the boundary, that occupation matching is not a prediction of ability or success, and the basis of the Five branches findings. The "Response quality" heading gives the readable quality label and its one-sentence explanation (ADR-044 §4) and explains the flags in plain language; if there is no flag, it says that none of the five indicators gave a signal and how many attention-check items the report contained. The "In your view" heading is for the feedback summary (5.8). The "Sources" heading links to the changes (`/analysis/changes`) and trust (`/analysis/trust`) pages; the appendix ends with a version stamp. The stamp text still reads "Form v4"; it should be aligned with form v5 (ADR-044).

None of the Turkish B2C products surveyed in the competitor products note publishes reliability figures. The missing piece is the Turkish version's own coefficients; they will be added with Gate 1 (Chapter 4).

5.3.5 Single vocabulary

The universal names of the report are kept in a single file: `Front-end/src/components/test/report/vocab.ts`. The axis letters and their English names (`AXIS_EN`), the axis pairs and the Big Five domain on which each rests (`AXIS_PAIRS`), the universal names of the temperaments (`TEMPERAMENT_EN`), the English names of the Holland areas (`HOLLAND_EN`), and the order and English names of the domains (`DOMAIN_ORDER`, `DOMAIN_EN`) are here. The equivalents in each language are under `Front-end/messages/<dil>/report.json` for five languages (`domain`, `temperament`, `axis`, `clarity`). The Turkish names of the Holland areas are the same in the front end and in Pusula-API (`InterestItem::NAMES`; ADR-031 §5). The rationale for the single vocabulary is the text-level counterpart of the "single source of a derived value" principle of ADR-016: if the same concept appears in two places under two different names, the reader may think that two different things were measured (inference).

5.4 Comparison and percentiles

5.4.1 The norm problem and the history of the decision

Every comparative statement such as "higher than 80% of your peers" requires a norm group. A raw score, by contrast, supports only two interpretations: a content interpretation that depends on the meaning of the items ("to these items you mostly answered 'like me'") and a within-person ranking ("your two highest interest areas"). Even saying that a dimension is "high" requires a point of comparison (scoring and norms note, S1 inferences).

For norm size the European standard is clear. For low-stakes use, EFPA regards fewer than 200 people per norm group as inadequate and 1000 or more as excellent (EFPA, 2025, §8.1.3). Regression-based norming requires a sample 2.5–5.5 times smaller for the same precision (Oosterhuis et al., 2016); cNORM reaches adequate norm quality with about 100 cases per cohort (Lenhard & Lenhard, 2021). In Hogan Personality Inventory data, on the other hand, $N > 100$ was found to have little practical effect on the reliability of standard scores, whereas **which norm group was chosen** made more of a difference than sample size: an average difference of 7.3 T-score points was observed between different occupational norms (Tett et al., 2009).

Rehnüma's answer to this question has changed three times. ADR-026 §6 provided only for raw scores and within-person ranking while there was no norm. ADR-032 (23 October 2026) redefined the percentile as "rank among those who have taken the test on Rehnüma" and displayed it once the reference group reached 30 people; the same decision had rejected Johnson's (2014) English norms as misleading because of the difference in language and culture. Version 1.1 had defended this choice and had also recorded the volatility of a small group: by the binomial approximation, in a group of 30 the percentile of the person in the middle fluctuates by roughly ± 18 points (inference).

ADR-082 (22 November 2026) reversed the decision: **personality percentiles come from the scale's own published data.** The rationale emerged in use. The reference group had reached 683 people, but it had not been selected at random and it contained trial runs; the sentence "higher than 10% of the 683 people who took it" did not inspire confidence. A large, documented sample that anyone can download was judged more defensible. This change trades one risk for another. Sampling error and the problem of a self-selected, contaminated group largely disappear: 619,150 people is far above EFPA's highest threshold. On the other hand, the ground on which ADR-032 rejected the norm has not been refuted: carrying the norm of the English form over to the Turkish adaptation is an assumption, and ADR-082 accepts this as a "known limit". By the finding of Tett and colleagues (2009), what makes the difference is which group it is rather than how large it is (5.4.4).

5.4.2 Computing the published norm

In Pusula-API the norm is read from the class `app/Profile/PublishedNorms.php`. The source is Johnson's (2014) open IPIP-NEO-120 data: 619,150 people (https://osf.io/wxvth/, `IPIP120.dat`). Because the published 120-item means do not fit Rehnüma's two stages, the norm was recomputed from the responses of the same people: 60 core items for stage 1 and 120 items for stage 2 are separate tables. For every scale, the mid-rank percentile of every possible total score is stored; no normal distribution is assumed. The person's score is looked up in the table; if it falls between two steps a linear interpolation is taken, and the result is kept between 1 and 99. If the facets have been completed to 4 items, the domain score is compared against the stage 2 table. The table (`database/data/norms_ipip_neo120.json`) and the script that generates it (`database/data/tools/build_ipip_norms.py`) are in the repository; the 95 MB of raw data is not.

There are two exceptions. The items were matched by text: 117 of the 120 items match their English original in the official key exactly and in the same direction; the three O6 items that were rewritten to replace originals based on United States politics are not in the original data. For this reason **no percentile is given for facet O6**, and the Openness to experience domain is normed with the remaining original items (11 of 12 items at stage 1, 21 of 24 at stage 2). Second, **there is no percentile for interest areas**: because no published norm table exists for the O*NET Interest Profiler, only the 0–100 score is shown (ADR-082 §4–5).

The `ReferenceGroup` class and the `norms_min_n` setting remain in the code but are not connected to the report; they will be reused if a Turkey-specific norm is collected in the future. The Rehnüma back end restricts the `norms` field to `n` and the domains in the summary and passes it in full in the full report (`PersonalityReport.php`). The percentiles of the sample report (`/analysis/report/sample`) were computed from the same table; the page says that the profile is not a real person but a synthetic profile (ADR-031 §2).

5.4.3 Language: naming the source

The percentile appears in two places: in the strip on the At a glance page and on the cards of the Five core traits page, side by side with the 0–100 score. The Version 1.1 sentence "higher than P% of the N people who took it on Rehnüma", and its softened form at the extremes, have been removed from the cards (commit fdc0ed1). The work of the rule for the extremes is done today by the table: the percentile does not fall below 1 and does not rise above 99.

How the source is to be stated is a separate user decision (ADR-082 §6): the text does not say "not a Turkish norm"; it states what the source is in affirmative wording: "personality percentiles are relative to the published international sample of the IPIP-NEO-120 (Johnson, 2014; 619,150 people)". Negative phrasings have been removed in the five languages (commit d018471). Today the sentence appears on the trust page and, in shortened form, on the entrance page of the analysis; the trust page additionally says "English form". The affirmative construction says nothing false and gives the norm group by name and by size (scoring and norms note, S1). Its price is that the reader has to draw for themselves, from the phrase "international sample", the conclusion "not relative to my peers in Turkey"; with respect to ITC (2005) guideline 29.1, this is a weaker disclosure than the explicit negative note in Version 1.1 (inference).

A second weakening is inside the report. ADR-082 had provided that the source would also be written in the note on the closing page; that disclaimer paragraph was removed the same day in the simplification (commit 286a3bb). Today no page of the report says against which sample the percentile is given; the item in the technical appendix, "Scores are on a linear 0–100 scale; they are not percentiles", describes only the score. To see the source, the reader has to go to the trust page.

5.4.4 Open issues

The published-norm implementation has five open points; the first three are ADR-082's own list of "known limits".

The first is **the equivalence assumption.** The norm comes from volunteers who completed the English form online. The psychometric validation of the Turkish adaptation has not been carried out, 34 items were rewritten in the translation review, and there is no evidence of measurement invariance between the two languages. Unless scalar invariance holds, it cannot be said that the same raw score indicates the same level in the two languages; this is precisely the ground on which ADR-032 rejected the foreign norm as "misleading". The assumption should be tested with Gate 2 and Gate 4 data (Chapter 4).

Second, **the norm does not distinguish age or gender**; the sample contains every age and country together. The mean Conscientiousness and Agreeableness levels of adolescents differ from those of adults (maturation literature; scoring and norms note, S7); the 14–17 age group may appear systematically low in these domains. This problem is treated separately in Chapter 8.7 under the heading of algorithmic fairness.

Third, **the international means lie above the midpoint of the scale**: in the norm table the domain means range from 53.4 to 65.0 at stage 1 and from 54.6 to 66.9 at stage 2. The same score can fall to a lower percentile than it did against the old reference group; the synthetic profile in the sample report shows 1% in four domains (ADR-082). The band of the domain sentence, however, is chosen by the raw score (60 and above high, 40 and below low; `ProfileBuilder.php`): the sentence of the "high" band and a percentile below the middle can appear side by side on the same card (inference).

Fourth, **the source and the uncertainty are not next to the number** (5.4.3): the source of the norm is not written in the report, and the measurement error in the person's own score is not shown (Table 5.2).

The fifth is the transitional state. According to ADR-082, until Pusula-API is deployed to production, live reports continue to be computed with the old group. If `ReferenceGroup` is reused, the shortcomings recorded in Version 1.1 (results with quality flags not being excluded from the group, and the absence of a separate n and threshold for facet percentiles) should also be remedied.

5.5 Universal codes and their Turkish equivalents

In its first version the report foregrounded the trait code and the temperament under Turkish names and kept the universal codes (ISTP, SP) in the background. ADR-032 §4 reversed this order: the codes are shown in the universal letters and with their English names, and the Turkish equivalent appears in the explanation. The rationale is that the person should be able to match a code they have read in other sources with the one given here. If a user has seen the result "ISTP" in another test, they should be able to read, in the same letters, which axes the four letters here describe. The vocabulary used is given in Table 5.6.

Table 5.6. The universal codes in the report and their Turkish equivalents.

Layer	Universal code and name	Turkish equivalent	Measurement it rests on
Trait code, 1st letter	E Extraversion / I Introversion	"Enerjinin kaynağı" (Source of energy)	Extraversion (E)
Trait code, 2nd letter	N Intuition / S Sensing	"Bilgiyi alma" (Taking in information)	Openness to experience (O)
Trait code, 3rd letter	F Feeling / T Thinking	"Karar verme" (Making decisions)	Agreeableness (A)
Trait code, 4th letter	J Judging / P Perceiving	"Dış dünyaya yaklaşım" (Approach to the outside world)	Conscientiousness (C)
Temperament	SP · Artisan	"Deneyerek öğrenen" (Learns by doing)	Low O + low C
Temperament	SJ · Guardian	"Düzen kuran" (Builds order)	Low O + high C
Temperament	NF · Idealist	"Anlam arayan" (Seeks meaning)	High O + high A
Temperament	NT · Rational	"Çözüm arayan" (Seeks solutions)	High O + low A
Interest code	R Realistic, I Investigative, A Artistic, S Social, E Enterprising, C Conventional	Turkish names of the Holland areas (e.g. "Gerçekçi", "Düzenli"; Realistic, Conventional)	O*NET Interest Profiler
Five domains	Openness, Conscientiousness, Extraversion, Agreeableness, Emotional stability	"Deneyime açıklık", "Sorumluluk", "Dışadönüklük", "Uyumluluk", "Duygusal denge" (Openness to experience, Conscientiousness, Extraversion, Agreeableness, Emotional stability)	IPIP-NEO-120

Note. Source: `vocab.ts` (`AXIS_EN`, `AXIS_PAIRS`, `TEMPERAMENT_EN`, `HOLLAND_EN`, `DOMAIN_EN`) and `messages/tr/report.json`. Emotional stability is the reverse direction of the Neuroticism scale in the original; the item content is the same (`/analysis/changes`).

The axis mapping rests on the finding of McCrae and Costa (1989): the four MBTI indices are related to four Big Five domains ($r \approx -.74$ between E–I and E, $r \approx .72$ between S–N and O, $r \approx .44$ between T–F and A, $r \approx -.49$ between J–P and C; values from a secondary citation). These correlations show why the mapping is defensible, but also why the axes are not equally dependable: the T–F and J–P let-

ters are the most fragile (derived layers note, §2). The same study found no support for "true dichotomous preferences"; the distributions are not bimodal. The letter is therefore the label of a continuous score and is given on every axis of the report together with a clarity label (5.1.3).

The temperament partition follows Keirsey's theoretical classification: the first split is between concrete and abstract (S/N), and the second split is on J/P on the concrete side and on T/F on the abstract side; the second axis of the two sides rests on different factors (derived layers note, §3). No evidence could be found that the four temperaments form an empirical cluster in Big Five space, nor did we come across a peer-reviewed study reporting direct correlations between the Keirsey temperaments and the NEO or the BFI. For this reason the temperament is only a narrative layer in the report; no suggestion on the dashboard changes because of it (ADR-033 §1; Chapter 6).

Foregrounding the universal codes carries a tension. Because the code uses the same letters as the MBTI, the reader may take it for an MBTI result. Three precautions had been taken. The first is the note saying that the code "is derived from Big Five scores" and "is not related to Myers-Briggs" (a mandatory note in ADR-026 §3). Pusula-API continues to produce this note together with the result, but the line on the motif page and the disclaimer paragraph in the closing were removed in the 22 November simplification (commit 286a3bb); today the note is read not in the report but on the trust page. This discrepancy should be resolved. Second, the texts of the 16 motifs and 4 temperaments were not taken from MBTI, Keirsey or 16Personalities texts but were written from the Big Five counterparts of the letter meanings; the motif names are also Rehnüma's own (ADR-028 §6). Third, the trait code and the temperament are in the full report only; they do not appear in the identity line of the summary (`pages/Overview.tsx`). A trademark opinion on the four-letter code is among the legal clearances that need to be obtained before launch (ADR-026, Impact; Chapter 8.4).

For the Holland code the situation is different. The RIASEC letters and their English names are O*NET's own terminology and match the occupational data one to one. Here the universal code enables the person to compare their result with O*NET occupation profiles. The Holland colours are used only in the dots and letter circles, as a pale ground; the name of the area is always given in writing (ADR-032 §3, ADR-080 §6). The Holland letters are written in ink colour (ADR-031 §7).

5.6 Visualisation decisions

The common principle of the charts is "one series, one accent" (ADR-032 §3; `report/charts.tsx`). Every chart shows a single data series; the colour palette consists of ink and a single accent colour. In Version 1.1 the accent was orange; the design system of the analysis side today uses olive as the single accent (ADR-031 update, ADR-080 §6). Colour does not carry identity: what an element is is always said in writing, and colour is used only to say "this is you" or "this is in the lead". This principle has two rationales. The first is accessibility: a reader with colour blindness never has to infer any information from colour alone. The second is semantic discipline: a single accent gathers attention on the person's position (inference). The chart types used, and what they deliberately do not show, are given in Table 5.7.

Table 5.7. Chart types in the report.

Chart	What it shows	What it deliberately does not show	Where it is used
Distribution curve (`Distribution`)	A normal approximation from the mean and deviation of the published norm, the mean (dashed), the person's line; shaded area = those below	The empirical distribution; a "good/bad" direction (pole names at the ends)	Five core traits
Bipolar axis (`Bipolar`)	Two letters, the middle line (50), notches at 25 and 75, an accent extending from the centre to the person	A one-directional "score bar"; the certainty of the letter (clarity is written separately)	Mind motif
Percentile strip (`PercentileStrip`)	A 0–100 band, the middle half (25–75) as a grey area, the person's dot	A percentile when there is no norm (falls back to the raw score)	At a glance (the five domains only)
Keirsey 2×2 map (`QuadMap`)	Horizontal O (S ↔ N); vertical C on the S side, A on the N side; the active cell accented	A claim that the temperaments are an empirical cluster	Temperament
Ranked bars (`RankBars`)	Rank number, label, value; the first element accented	Colour identity for the categories; a percentile for interest areas (ADR-082 §5)	Types, interests, tendencies, values
Similarity rings (`ArcRank`)	The percentage of the four closest types, concentric arcs	A pie slice (a part–whole reading)	At a glance, full report
Holland hexagon (`HollandHexagon`)	The shape of the six areas in the adjacency order of the model	The independence of the areas	Interests, summary

Note. Source: `Front-end/src/components/test/report/charts.tsx`, `test/HollandHexagon.tsx`.

The distribution curve is a normal density curve drawn from the mean and standard deviation in the norm table. The person's score is marked with a line in the accent colour, and the area beneath it is lightly tinted. Two pole names are written under the curve (such as "Likes the familiar" and "Seeks the new"); the use of pole names instead of 0 and 100 prevents a "high = good" reading. One limitation of this chart should be recorded explicitly: whereas the percentile is read from the empirical rank in the table, the curve is a normal approximation. If the distribution is skewed (e.g. a ceiling effect in Agreeableness), the shaded area under the curve and the percentile written out may diverge from each other (inference). Because the table carries the percentile of every possible score, the curve could be drawn from the empirical distribution; this has not been done. When there is no norm the curve is not drawn; only the person's line and the axis remain.

The bipolar axis shows the four letters of the trait code. The accent bar starts not from zero but from the centre and extends to the person's score; the length of the bar shows not "how much" but "how far from the centre", that is, clarity. The design makes the letter read as a position on a continuous axis and is the visual counterpart of McCrae and Costa's (1989) critique of the dichotomy.

The percentile strip shows the five domains at a single glance. The grey area marks the middle half (from the 25th to the 75th percentile); if the dot is inside this area, the person is in the middle of the norm sample. This area keeps the reader from exaggerating small differences (inference).

The Keirsey map makes the logic by which the temperament is derived explicit. The horizontal axis is Openness to experience; the vertical axis is Conscientiousness on the concrete side ($O < 50$) and Agreeableness on the abstract side. This asymmetry, touched on in 5.5, is visible in the structure of the chart and is not hidden from the reader.

The similarity rings show the four closest of the eight types as concentric arcs. A pie chart was deliberately not used: although the percentages sum to 100 (ADR-029 §2), a pie slice suggests that the person is made up of parts, whereas the percentage is a relative expression of similarity to the types (inference).

The accessibility rules of the charts come from ADR-031 §7. The accent colour is not used directly as a text colour; numbers that need to be written with emphasis carry a separate accessible tone (`--accent-text`). The smallest type in the report is 13 pixels (ADR-080 §3). The SVG charts are marked `aria-hidden` and the same information is given in writing beside them. Elements with no written counterpart, such as the bipolar axis and the percentile strip, carry `role="img"` and a readable `aria-label` (e.g. "Introversion 38, Extraversion 62; Moderate").

Motion is gathered in a single hook with GSAP (`report/motion.ts`, `useReportMotion`; ADR-032 §5). Sections appear by rising 14 pixels, bars extend with a `scaleX` transform, numbers count up from zero to their values, markers slide to their targets, and curves are drawn. The page transition is 260 milliseconds. Motion is done through transform and opacity, not size. Under the reduced-motion preference (`prefers-reduced-motion: reduce`) only opacity changes; numbers and bars appear directly in their final state, and the page transition drops to a 150-millisecond fade. Motion may also have a cost for reading the report: a counting number can lead the reader to see one of the intermediate values and remember it wrongly (inference). For this reason motion plays only once, when the section opens; in subsequent interactions the values are fixed.

5.7 Trust design

5.7.1 The source of trust: design and fact

ADR-031 records the user's goal as follows: the person should want to take the test while taking it and should feel its scientific trustworthiness as design. The decision is that trust comes not from marketing but from design and fact (ADR-031 §2). On the entrance page of the analysis the instrument names and item counts, the sources and the licence notice, the limits ("The Turkish adaptation is in its pilot stage; personality percentiles are relative to the scale's published international sample") and what happens to the data ("If you don't sign up, they are deleted after 30 days") are visible. There are no marketing clichés, no invented social proof and no urgency. The sample report is a real Pusula-API output, and the page states that it is a synthetic profile.

The most comprehensive application of this principle is the public trust page (`/analysis/trust`; `Front-end/src/app/[locale]/(marketing)/analysis/trust/page.tsx`, text in five languages in `trust.json`). The page opens with four fact cards (the original sample of the IPIP-NEO-120, 30 facets, 910 active occupations, and the median facet recovery of r = .89 at Gate 0) and continues with seven sections: what is measured, what it rests on (source, known reliability, licence), how it was tested (the five gates and their status), the principles of honesty, what is not claimed, how the result is reflected on the dashboard, and what happens to the data. An eighth section links to this monograph.

The "What we do not claim" section says the following: the Turkish version has not yet been validated by an independent study; personality percentiles are relative to the published international sample (English form) and no percentile is given for interest areas; the instrument is not a diagnostic tool and does not replace a clinical assessment, career counselling or a hiring decision; occupational fit is a similarity of interests and working style, not a prediction of ability or success; the mind motif and the temperament are narrative, and no suggestion on the dashboard changes because of them. This list largely coincides with the limitations that ITC (2005) guideline 29.1 requires to be disclosed to the user.

The trust page's own consistency also has to be checked. The two discrepancies recorded in Version 1.1 have been resolved: the page explains the number of occupations as "910 active out of 923 records" and shows Gate 1 as "Pending". A new discrepancy has arisen in their place: the list of principles of honesty still says "Locked sections do not show blurred or fake content; the table of contents says what will open"; after ADR-081 this sentence is not true (5.7.2). A false sentence on a page that claims to rest on fact loses more trust than the same error on a marketing page (inference).

5.7.2 Displaying locked content: from a table of contents to a blurred continuation

In moving from the free view to the full report the product has to withhold something; the question is how what is withheld is to be displayed. Rehnüma has given two different answers to this question. Under ADR-031 §4 and ADR-032 §1, the page of a locked section showed only a table-of-contents sentence and the "Open the full report" box; ADR-031 had defined this explicitly as "instead of blurred lines", and Version 1.1 had set this choice against the dominant pattern in the market. ADR-081 (22 November 2026) changed the decision, because the person was being left to a payment decision without seeing anything about their own result. Today the locked section opens (`pages/LockedPage.tsx`): at the top is the introductory sentence of the section, beneath it a real opening sentence from the person's own result, and beneath that a blurred continuation and the lock box. The blurred part is not real report text; it is a filler built from the introductory sentences of the other locked sections, it cannot be selected and it is hidden from screen readers. In the Occupations section, what is shown blurred is repeated rows of the three occupations that already arrive in the summary. The Five branches section carries no opening sentence.

For comparison we took the dominant pattern in the market. 16Personalities frames its free result as "about 5% of what we can show" (16Personalities, n.d.-b; competitor products note); in our observation of 23 October 2026, its locked sections were seen to show an unreadable fake text with scrambled letters. The comparison of the two strategies is in Table 5.8.

Table 5.8. Two strategies for displaying locked content.

Criterion	Scrambled fake text (16Personalities, observation of 23 October 2026)	Real opening sentence and blurred filler (Rehnüma, ADR-081)
Information given to the reader	The impression that content exists and is long	What the section is about, one sentence from the person's own result, and the impression that a continuation exists
Link to the measurement	The visible texture is not tied to the person's score	The opening sentence is from the person's result; the blurred part is not tied to the score
Data sent from server to client	Fake texture; the real text is not sent (inference)	Only the first sentence per section (`PersonalityReport::summary`, `openings`)

Criterion	Scrambled fake text (16Personalities, observation of 23 October 2026)	Real opening sentence and blurred filler (Rehnüma, ADR-081)
Psychological mechanism	Curiosity gap and a volume signal (inference)	Informing by example together with a curiosity gap (inference)
Relation to the Barnum principle	An image not tied to a score presented as "specific to you" (inference)	The blurred filler can be read as the continuation of the person's own text (inference)
Known user reaction	Complaints about the unexpected wall (Trustpilot 2.5/5; competitor products note)	Not measured

Note. The 16Personalities observation is a screen inspection on a single date; the practice may differ on other pages of the product or at other times. The "Psychological mechanism" row is inference.

Version 1.1 had criticised the scrambled-text technique as follows: the reader thinks they are seeing a part of their own report in blurred form, whereas what they see is tied to no measured score, and this is the visual counterpart of the Barnum principle. This criticism applies today to Rehnüma's blurred filler as well. The differences that remain are these: the readable sentence is real and comes from the person's own result; the paid text is never sent to the browser. That the filler is not real text, however, is not told to the reader. The labour illusion literature explains the mechanism: visible effort and volume can raise perceived value, but increased dissatisfaction when an unsuitable result was shown (Buell & Norton, 2011).

The decision is a shift in values: ADR-081 gave priority to the person seeing something of their own result before the payment decision. The effect of the two designs on conversion and on trust has not been measured. ADR-031 §4(b) and the corresponding principle on the trust page should be updated in line with the new decision; its effect on the commercial model is discussed in Chapter 8.6.

5.8 The feedback loop

In the motif, temperament, archetype, strength and three life-area sections of the full report there are two buttons: "Fits" and "Doesn't fit" (`report/parts.tsx`, `Fits`; ADR-028 §5). The buttons appear only in the full report of a signed-in user; they are absent from the summary and from the sample report. The choice is shown immediately, the request goes in the background, and the person is told "we will use this to improve the report". A new reaction given to the same element overwrites the old one.

ADR-044 §5 requires the reactions to be read back to the person. `PersonalityTestService::feedbackSummary` collects, per section, the "fits" and total counts and the elements marked "doesn't fit"; the "In your view" heading in the technical appendix writes this as "You reacted to N items; you said M fit you". In the current code the summary is added only on the path that serves the report through the bridge record (`report`); the `profile` method used by the interface does not pass the summary, and the heading is not visible today. This is a wiring gap.

The data are stored in the Rehnüma back end in the `report_feedback` table (`Back-end/database/migrations/ 2026_10_23_200000_create_report_feedback_table.php`). Each row carries the user, the section of the report (e.g. `motif`), the element (e.g. `INFJ.work`), the reaction (`fits`) and a context field. The context is filled by `PersonalityTestService::feedback` with the trait code, temperament, closest archetype and test stage at the moment of the reaction. The function of this snapshot is to find

the opposite profile in the pruning analysis: even if the person later retakes the test and the label changes, the label under which the reaction was given is not lost. The triple of user, section and element is unique. The endpoint is limited to 60 requests per minute.

The use of the feedback is tied to two decision rules. Gate 4 of ADR-029 provides for a sentence whose "doesn't fit" rate is 40% or above to be removed from the report. ADR-028 §5 adds the discrimination criterion: a sentence that everyone marks "fits", including those with the opposite profile, is not discriminating. The second rule is the empirical form of Poškus et al.'s (2014) proposal for statement-level evaluation and of the base-rate test in the scoring note. The list of these two rules and of our proposals that complement them is in Table 5.9.

Table 5.9. Decision rules for the feedback data.

Rule	Criterion	Decision	Basis
Misfit pruning	"Doesn't fit" ≥ 40% for a sentence	The sentence is removed or rewritten	ADR-029 Gate 4
Discrimination	The "fits" rate among those with the opposite profile is close to that among those with the target profile	The sentence is not counted as discriminating and is rewritten	ADR-028 §5; Poškus et al. (2014)
One-way filter	A high "fits" rate	A sentence never gains validity from this rate; only a low rate is a ground for elimination	5.1.1; Poškus et al. (2014)
Comparison within valence	Cost and strength sentences are compared separately	Negative sentences are not eliminated disproportionately because of positivity bias	Dickson and Kelly (1985); inference
Minimum number of reactions	Sufficient reactions per sentence (threshold not set)	No decision is made below the threshold	Inference; Gate 4 analysis plan

Note. The first two rows have been decided in ADRs; the last three rows are the proposal of this chapter and have not yet been written into an ADR.

The third row of the table is the basic principle of the loop: because the Barnum literature shows that high acceptance is not evidence of validity, sentences are eliminated as failures and are not confirmed as successes. The fourth row is a precaution against a side effect. Given that positive sentences are known to be accepted more easily (Dickson & Kelly, 1985), a single threshold eliminates cost sentences more often than strength sentences, and over time the report could turn back into a purely positive text. This would be the erosion of Rule 3 (the cost–benefit pair) by way of feedback. Applying the threshold separately according to the valence of the sentence reduces this risk (inference).

The loop also has known limitations. The first is selection bias: who presses the button is not random. The second is the scope of the context snapshot: the context stores the labels (code, temperament, archetype) and does not store the facet scores; if the person completes stage 2 later, the score at the moment of the reaction is lost. We recommend that the facet bands also be added to the context. The third is that the analysis tool does not yet exist: the only tool of the validation analysis, `docs/research/scripts/psychometrics.py`, processes the response matrix and does not process the feedback data. This analysis has to be written by Gate 4.

That the report texts have not been validated is explicitly on record (the note in `report_texts.json`); this loop is a way not of validating them but of finding the weakest among them.

Chapter references

- 16Personalities. (n.d.-b). *Premium*. https://www.16personalities.com/premium
- American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. https://www.ucc.ie/en/media/academic/appliedpsychology/ethicsforms/apa_code2002.pdf
- Buell, R. W., & Norton, M. I. (2011). The labor illusion: How operational transparency increases perceived value. *Management Science, 57*(9), 1564–1579. https://doi.org/10.1287/mnsc.1110.1376
- Chauliac, M., Willems, J., Gijbels, D., & Donche, V. (2023). The prevalence of careless response behaviour and its consequences on data quality in self-report questionnaires on student learning. *Frontiers in Education, 8*, 1197324. https://doi.org/10.3389/feduc.2023.1197324
- Dickson, D. H., & Kelly, I. W. (1985). The "Barnum effect" in personality assessment: A review of the literature. *Psychological Reports, 57*(2), 367–382. https://doi.org/10.2466/pr0.1985.57.2.367
- EFPA. (2025). *EFPA review model for the description and evaluation of psychological and educational tests* (Version 2025.5). European Federation of Psychologists' Associations. https://www.infocop.es/wp-content/uploads/2025/09/efpa_test_review_model_v2025_5.pdf
- Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology, 44*(1), 118–123. https://doi.org/10.1037/h0059240
- International Test Commission. (2005). *International guidelines on computer-based and internet-delivered testing*. https://www.intestcom.org/files/guideline_computer_based_testing.pdf
- Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003
- Lenhard, W., & Lenhard, A. (2021). Improvement of norm score quality via regression-based continuous norming. *Educational and Psychological Measurement, 81*(2), 229–261. https://doi.org/10.1177/0013164420928457
- Low, K. S. D., Yoon, M., Roberts, B. W., & Rounds, J. (2005). The stability of vocational interests from early adolescence to middle adulthood: A quantitative review of longitudinal studies. *Psychological Bulletin, 131*(5), 713–737. https://doi.org/10.1037/0033-2909.131.5.713
- McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality, 57*(1), 17–40. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x
- Oklahoma Baptist University. (2023). [Perceived accuracy of general and specific test results, Scholars Day poster]. https://scholarlycommons.obu.edu/scholars_day_conference/2023/posters/38 (reference to be verified)
- Oosterhuis, H. E. M., van der Ark, L. A., & Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment, 23*(2), 191–202. https://doi.org/10.1177/1073191115580638
- Poškus, M. S., Kairys, A., Liniauskaitė, A., & Žukauskienė, R. (2014). "Your NEO summary": Is it a valid way of providing feedback, or is its accuracy due to the Barnum effect? *Psichologija, 49*, 44–59. https://doi.org/10.15388/Psichol.2014.49.3694
- Pullmann, H., Raudsepp, L., & Allik, J. (2006). Stability and change in adolescents' personality: A longitudinal study. *European Journal of Personality, 20*(6), 447–459. https://doi.org/10.1002/per.611

- Studia Doctoralia. (n.d.). [Visual format of IPIP feedback and perceived accuracy]. https://journals.unibuc.ro/index.php/sd/article/view/1300 (reference to be verified)
- Tett, R. P., Fitzke, J. R., Wadlington, P. L., Davies, S. A., Anderson, M. G., & Foster, J. (2009). The use of personality test norms in work settings: Effects of sample size and relevance. *Journal of Occupational and Organizational Psychology, 82*(3), 639–659. https://doi.org/10.1348/096317908X336159
- University of Oxford. (n.d.). False personality feedback and mood [Thesis]. https://ora.ox.ac.uk/objects/uuid:c88e03b3-19ef-4929-9b06-add78422de06 (reference to be verified)
- University of Warwick. (n.d.). The Barnum effect and behavioural tendencies [Doctoral thesis]. https://wrap.warwick.ac.uk/id/eprint/96198 (reference to be verified)
- Zelenak, M. F. (2016). *2014 American Community Survey (ACS) Internet test* (American Community Survey Research and Evaluation Report Memorandum Series No. ACS16-RER-03). U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2016/acs/2016_Zelenak_01.pdf

6. Character Trace: Carrying Personality Findings into the Everyday-Life Dashboard

The preceding chapters dealt with how the test measures (Chapter 3), which gates that measurement must pass before it is considered trustworthy (Chapter 4) and how the result is explained to the user (Chapter 5). This chapter examines the question that begins where the report ends: what can a personality profile change, and on what evidence, in the goal, health and finance surfaces the user looks at every day? In Rehnüma this transfer is called the **character trace**. The main argument of the chapter is that the bulk of the transfer should consist of robust principles that are *independent* of personality, and that the layer that *depends* on personality should remain a narrow layer that only softens and can be undone. This argument rests on an inventory study (§6.2), an evidence-map review with six strands (§6.3–6.5), an evidence policy (§6.6) and a rule catalogue and engine that make this policy executable (§6.7). The chapter closes with a case that traces the output of the engine on the product's sample report profile (§6.8) and with an account of the open gaps (§6.9).

6.1 Problem and approach

6.1.1 The promise of ADR-025 §7

The decision to add the personality test to Rehnüma (ADR-025) did not design the test merely as a report generator. The seventh clause of the decision defines the place of the test within the product: a widget named "Pusulan" ("Your Compass") on the goals page, a `personality_nudge` type to be added to the `insights.type` field, and example rules for it. The examples fell under four headings: for low Conscientiousness, a smaller key-result step and frequent reminders; for low Emotional stability, gentle language in the budget against the ostrich effect (ADR-024); for high Openness to experience, variety in workouts; and sleep and workout time suggestions according to chronotype (ADR-025 §7).

These examples are intuitively reasonable; they are of the kind frequently met in the marketing language of personal development applications. Each of them, however, contains a testable claim. The sentence "a user low in Conscientiousness benefits from frequent reminders" assumes that the effect of reminder frequency is moderated by Conscientiousness. The sentence "gentle language for those low in Emotional stability" assumes that the effect of feedback tone interacts with neuroticism. If these assumptions are correct, the product helps the user by speaking according to personality; if they are wrong, it either adds an ineffective complexity or, worse, hides behind a label and presents the user with a wrong plan. Because the psychometric validation of the Turkish version of the test is continuing with the pilot, and because simulation has shown that the derived labels can change in 43–55% of cases at retest (Chapter 4; ADR-029), the second risk is not abstract.

6.1.2 Personalisation traps

The literature and product practice suggest that there are four distinct traps in personalising by personality. Naming them at the outset makes explicit the rationale for the policy decisions in the rest of the chapter.

First trap: confusing perception with behaviour. The great majority of studies examining the relationship between personality and persuasion strategy ask the participant how persuasive a strategy *appears* (Oyibo et al., 2017; Alkış & Taşkaya Temizel, 2015; Ndulue et al., 2022). The small number of studies that measure actual behaviour report that what people find more persuasive is not what moves them to act (Ciocarlan et al., 2019). The scoping review of personalised persuasive systems likewise finds no randomised controlled trial (RCT) showing that an intervention tailored to personality changes the behavioural outcome more than an untailored one (Alslaity et al., 2023).

Second trap: mistaking a main effect for an interaction. It has been strongly shown that low Emotional stability is associated with more anxiety, more emotional eating and more financial distress. This is a *main effect*: it describes the average condition of people in the low `S` band. To conclude from this that "a different language works better for a person in the low `S` band" is a claim of a language × personality *interaction*, and it requires separate evidence. As will be seen in §6.3, this interaction has been tested almost nowhere.

Third trap: overstating the stability of derived labels. Layers such as the mind motif, temperament and archetype are derived from Big Five scores and are categorical. A categorical label changes with a small measurement error if the score is close to the boundary; in the MBTI tradition it has been reported that 39–76% of types change at a five-week retest (Pittenger, 2005). Product behaviour tied to such a label carries measurement error directly into the user's plan.

Fourth trap: protecting by hiding information. Withholding negative information in order to "protect" an anxious user looks attractive. Yet the finance literature shows that making information easier to access reduces late fees by 38.4% (Carlin et al., 2023) and that negative relative information reduces spending by 12% (D'Acunto et al., 2019). No study in which hiding information helped appears in the evidence map. Personalisation should be done not in visibility but in tone, order and the suggested action.

6.1.3 Inventory first, then evidence, then rule

Against these traps we divided the work into three consecutive stages and made it a rule that no stage is entered before the previous one is finished.

1. **Inventory.** First, without writing any personality content, the code of every widget was read and a count was made of the states the user can be in and of the places where the widget says, chooses or suggests something. At this stage no sentence of the kind "extraverts forget to drink water" was written; possible effects were recorded only as research questions (`Front-end/src/widgets/KARAKTER-IZI.md`, "Kurallar").
2. **Evidence.** The scenarios produced by the inventory were distributed across six research strands; each strand produced a sourced note, and these notes were combined in a binding synthesis report (`docs/research/reports/Karakter izi kanıt haritası.md`).
3. **Rule.** Only after this was a status assigned to each scenario according to its evidence grade, and the records were written to a machine-readable catalogue; the catalogue contract was protected by tests (ADR-033).

The purpose of this ordering is to make it structurally difficult to put intuition in the place of evidence when writing rules. The inventory derives from the code the list of questions for which evidence is to be sought; the evidence map gives each question a grade; the rule catalogue turns into behaviour only

those questions whose grade suffices. A question whose evidence does not suffice is not deleted; it enters the catalogue as a "no change" record together with its rationale. In this way, why a scenario does not differ by personality becomes as traceable as why it does.

6.2 Inventory method

6.2.1 Scope: fifteen surfaces

The unit of the inventory is the "surface": every part of the product that says, chooses or suggests something to the user and that personality information could in principle affect. The inventory was carried out on 23 October 2026, on the product as it stood that day; the list of surfaces and the counts in this subsection belong to that date, and later changes are given in §6.2.8. The grid catalogue (`Backend/config/widgets.php`) contained thirteen widgets at that date: in the goals module, Quarter Focus (`quarterFocus`); in the health module, Body Scan (`bodyScan`), Weight Tracking (`weightTrend`), Energy Needs (`energyNeeds`), Nutrition Plan (`nutritionPlan`), Water Tracking (`waterTracker`), Steps (`stepCounter`), Workout Programme (`workoutPlan`), Sleep (`sleepPlan`) and Pill Box (`pillBox`); in the finance module, Budget (`budget`), Vault (`vault`) and Market (`market`). Two surfaces were added to these. The first is phone sync (`health/sync`); it is not a widget but the source of step and sleep data, yet it has decision points of its own, such as automation times and the stale-data warning. The second is the cross-cutting surfaces that appear on every page: celebration (ADR-010) and the notification bell and service (ADR-019) were treated in a single file (`Frontend/src/components/app/KARAKTER.md`), because the celebration and reminder scenarios of the widgets are in practice implemented in these shared components.

The inventory also recorded four surfaces that did not exist at that date but matter for the character trace, though it wrote no file for them: the "Pusulan" widget (the visible place of the profile on the dashboard), the habit check-in list (`api.habits` and `api.checkIn` exist in the back end, there is no surface in the front end), the dashboard and `insights` cards (the table exists, it has no consumer, and the `personality_nudge` type is not in the enum) and the Tasks module (no widget in the catalogue). Because it was envisaged that tracking on the goals page would be done from there, the habit check-in scenarios were taken into the Quarter Focus file as additional decision points (H1–H7) (`goals/quarterFocus/KARAKTER.md` §3.1; the file was removed with ADR-035 and remains in the repository history).

6.2.2 Shared vocabulary

The first product of the inventory is the fixing of the variables produced by Pusula-API in a single vocabulary. The names in the vocabulary correspond one to one with the `database/data/report_texts.json` file of Pusula-API, and throughout the inventory the rule "no other name is invented" was applied. The vocabulary also shows the report tier at which each variable is available (ADR-025 §6): the summary profile (stage 1) contains only the domains, interests and a provisional archetype; the facets, motif, temperament, values and rhythm require the full report (stage 2). This distinction is the basis of the tier gate in §6.6. Table 6.1 summarises the vocabulary.

Table 6.1. Vocabulary of Pusula variables and report tier.

Layer	Values	Tier
Domains (Big Five)	`O` Openness to experience · `C` Conscientiousness · `E` Extraversion · `A` Agreeableness · `S` Emotional stability (high = stable)	summary
Facets (30)	Six facets for each domain; e.g. S1 Managing worry, S3 Mood balance, S5 Impulse control, C2 Orderliness, C4 Achievement striving, C5 Self-discipline, C6 Deliberation, O5 Curiosity about ideas, E6 Cheerfulness	full
Mind motif	Four axes (E/I, S/N, T/F, J/P), 16 codes	full
Temperament (Keirsey)	`SP` · `SJ` · `NF` · `NT`	full
Self type	8 archetypes, % similarity	summary (1), full (8)
Natural tendencies	10 tendencies (e.g. Analytical thinking, Planning and order)	full
Work values	6 values, ranked (Achievement, Independence, Recognition, Relationships, Support, Working conditions)	full
Rhythm (chronotype)	Morning type · Intermediate type · Evening type	full
Interests (RIASEC)	Three-letter code and distinctness	summary
Quality	`ok` · `caution` · `uninterpretable`	summary

Note. Source: `Front-end/src/widgets/KARAKTER-IZI.md`, "Pusula'nın verdiği değişkenler". Domains are converted to a `high/mid/low` band and facets to a percentile; at the inventory date the source of the percentile was Rehnüma's own reference group (ADR-032), and since 22 November 2026 it has been the published norm of the IPIP-NEO-120 (ADR-082). The Quality column gives the three names in the inventory vocabulary; the mapping of Pusula-API's four levels onto these is in §6.6.2. "Neuroticism" in the literature is read in this work as low `S`.

Two design decisions in the vocabulary shaped the later stages. The first is the reverse coding of the `S` axis: because Pusula reports a high score as "stable", findings on high neuroticism in the literature were translated in every note as "low `S`"; each of the six notes explicitly repeats this translation in its scope paragraph. The second is the acceptance from the outset that widgets would work with bands and not with raw scores. This is the first step in limiting how far measurement error is reflected in the product; treating the band as "neck-and-neck" in the 40–60 percentile range in §6.6 is the continuation of this decision.

6.2.3 Template

For every surface a `KARAKTER.md` was written with the same six-section template. The sections of the template are as follows: (1) *What it does*: the function of the widget, its files, catalogue variants, lock requirements and related ADRs; (2) *User states*: each independent state axis, its values and where it is read in the code; (3) *Decision points*: every place where the widget says, chooses or suggests something, with the line of code; (4) *Character trace orientations*: a table of decision point × user state × the Pusula variable that affects it × expected type of difference × research question; (5) *Matrix to be filled in*: scenario rows and columns for temperament, domain band and rhythm; (6) *Count and open questions*. The fifth section was deliberately left empty at the inventory stage and the note "to be filled in

after in-depth research" was put in its heading; in this way the template formally prevented content from being written before the evidence arrived. After the evidence map was completed, every cell of these matrices was filled in by linking it to a record identifier in the catalogue (ADR-033 §5).

Two general rules accompanied the template. For a user whose quality is `uninterpretable` or who has not finished the test, the present behaviour of every widget remains exactly as it is, and this is written in every file as the "base state". In addition, the health page shows a plan and holds no logging; no scenario touches the protected 3D body component (`bodyScan/body/`).

6.2.4 Deriving user states from the code

The methodological core of the inventory is that user states are read from the code and not imagined. For each state axis, where it comes from in the code was written with the line number; a state that has no counterpart in the code but is required by the scenario was explicitly marked as "implicit / missing data". The value of this rule can be seen in the goal direction axis of the Weight Tracking file. In the code, direction is derived from whether the target is smaller than the *first weigh-in* (`insights.ts:39`); the energy goal (`energy_goal`) and the starting value of the key result are not consulted. If the target equals the first weigh-in, the user is treated as "gaining" and falls at once into the "reached" state. So although the state "I want to maintain my weight" appears to exist in the language of the product, it does not exist in the code; the inventory recorded it as an implicit state (`health/weightTrend/KARAKTER.md` §2.B). Similarly, the concept of a plateau is not named in the code; the combination "flat rate + target not reached" is an implicit plateau. The *frequency* of weighing is not measured either; only the day gap enters the weight of the exponential average.

The same approach produced similar findings in the other files. The Budget file counts fourteen state axes (setup, income source and regularity, envelope allocation, position within the month, sign of the safe-to-spend amount, spending pace band, envelope fullness, running a deficit/saving, logging frequency, use of recurring items, reflection status, category weight, size variant); among these, the "running a deficit" state is not named in the code and is defined implicitly by the combination of `safe_to_spend < 0` and `saved = 0` (`finance/budget/KARAKTER.md` §2.9). In the Sleep file the chronotype can be inferred only from the plan times; the existence of actual night data was recorded as a separate axis, because the applicability of the conflict rule in §6.6 ("measured behaviour overrides Rhythm") depends on this axis.

When the axis values are multiplied, each file also gave the number of "meaningful core combinations"; impossible or trivial combinations were weeded out with a rationale. For example, the full product in the budget is 576, while the meaningful combinations were reported as about 24, as 45 in Weight Tracking, about 40 in Quarter Focus and about 18 in sleep (36 with night data). These numbers are not the unit of the research; the unit is the scenario, defined below.

6.2.5 Decision points and scenarios

A decision point is every place where the widget says, chooses or suggests something to the user: a default value, suggestion copy, a reminder, a celebration, a warning, a choice of chart, an empty state, an error message. Each decision point was numbered with the prefix `K` and recorded with its place in the code. In Weight Tracking, for example, the log confirmation (K4) says "X kg recorded · ±Y kg compared with the previous weigh-in" and calculates the difference from the raw weigh-in; the header

(K6) shows the raw weigh-in as the main number; insight 1 (K12) adds, on rapid loss, the note "A bit fast. Staying under 1% of body weight per week prevents muscle loss." (`health/weightTrend/KARAKTER.md` §3). Effects triggered outside the widget (the weigh-in causing the key result to be recalculated, the goal-completion celebration) were also counted as separate decision points.

The **scenario** is the unit of the research: one decision point × one user state × the Pusula variables that could affect that decision. For each scenario the expected type of difference (language, frequency, default, visual, order) and a research question were written. The research question column was the only place where a hypothesis could be written at the inventory stage, and it had to be written in the interrogative. Table 6.2 shows four of the fourteen scenarios of Weight Tracking as examples.

Table 6.2. Example scenarios from an inventory file (Weight Tracking).

#	Decision point	User state	Pusula variable	Type of difference	Research question (summary)
S1	K6 main number (raw weigh-in), K7 direction	rate band: falling, rising, flat	S1, S3, `S` band	visual, language	In low `S`, does putting the trend forward as the main number reduce the reaction to daily fluctuation?
S2	K4 log confirmation (raw difference)	every weigh-in	S1, S3, E6	language, frequency	In which profile is showing the momentary difference discouraging, and in which is it information?
S5	K13 "X kg to go"	flat or in the opposite direction (plateau)	S3, C5, NF	language	On a plateau, which language reduces dropout in which profile? Should the "plateau" be named?
S9	K9/K10 chart default	data maturity	O5, Analytical thinking, NT	default, visual	Do data-oriented profiles want a longer range and the raw points brought forward?

Note. Source: `Front-end/src/widgets/health/weightTrend/KARAKTER.md` §4. At the inventory stage variables such as temperament and natural tendency were also listed as "may affect"; that they cannot be used as conditions was decided in the evidence map (§6.4).

The last column of Table 6.2 shows why the inventory has to come before the evidence: in S5 and S9 the temperament (NF, NT) and natural tendency candidates are listed honestly, because at the inventory stage it was not known which variable had evidence behind it. Narrowing the scenario from the outset to "Big Five only" would have left undocumented why the temperament layer was rejected. Every candidate variable listed in these files received an explicit acceptance or rejection in the evidence map.

6.2.6 Count

Table 6.3 gives the count of the fifteen files dated 23 October 2026. In total, 156 state axes, 310 decision points and 229 scenarios were extracted.

Table 6.3. Character trace inventory (23 October 2026): number of state axes, decision points and scenarios per surface.

Surface	State axes	Decision points	Scenarios
goals/quarterFocus	12	22 (15 widget + 7 check-in)	22
health/bodyScan	12	16	15
health/weightTrend	8	21	14

Surface	State axes	Decision points	Scenarios
health/energyNeeds	11	20	14
health/nutritionPlan	17	22	18
health/waterTracker	9	25	16
health/stepCounter	8	24	15
health/workoutPlan	11	29	18
health/sleepPlan	7	15	12
health/pillBox	8	22	14
health/sync	6	12	8
finance/budget	14	28	20
finance/vault	12	21	15
finance/market	11	20	16
components/app (celebration + notification)	10	13	12
Total	**156**	**310**	**229**

Note. Source: the 23 October 2026 version of the file `Front-end/src/widgets/KARAKTER-IZI.md`, "Sayım (2026-10-23)" (`1b30640` in the repository history). The present version of the file carries seventeen rows under the heading "Sayım (2026-10-24)" (§6.2.8). A state axis is an independent dimension in which the user can find themselves in a widget; the numbers of core combinations obtained by multiplying the axis values are given separately in the files.

The scenarios are distributed unevenly by type of difference. The great majority fall into three types: **language** (warning, celebration, empty state, suggestion sentence), **default** (setup preselection: envelope ratio, sleep time, target rate, reminder on/off) and **frequency/timing** (reminder, repetition of celebration, comparison period). **Visual/order** scenarios (which chart comes first, list order) are fewer. This distribution set the priority of the research: the literature on personality × feedback language and on personality × reminders was put first. About 60 of the frequency and timing scenarios depend on the notification channel, which even today writes only medication reminders; this is an infrastructure precondition that has to be resolved first whatever the evidence says.

6.2.7 By-product of the inventory: seven recurring gaps

Reading the code through the lens of personality also made visible product deficiencies that have nothing to do with personality. The inventory counted seven gaps that recur across all the files: (1) the notification service (`NotificationService`) writes only the `medication` type, that is, there are no reminders for water, steps, sleep, workouts, weigh-ins and the budget (despite the promise of ADR-019 §6); (2) there is no chronotype input, and the nutrition template, the sleep default, the workout time and the sync automation times are fixed; (3) the setting of goal direction and phase is scattered (the weight direction is derived from the first weigh-in, and the energy goal and the weight widget are not linked); (4) the "looking" event is measured nowhere, although the ostrich-effect scenarios depend on it; (5) there is no rule engine and no `personality_nudge` type; (6) the celebration and notification copy exists as duplicates in two places (`Celebration.tsx` and `NotificationService`); (7) deficiencies independent of character: the `SyncConnect` component is rendered from nowhere, sync freshness is not shown, the goals page does not show goals beyond pagination, and the `on_plan` sleep data are not used (`Front-end/src/widgets/KARAKTER-IZI.md`, "Bütün dosyalarda tekrar eden boşluklar"). Of

these, the rule-engine part of the fifth item was closed by the implementation of this chapter described in §6.7 (the `personality_nudge` type was not added to the `insights` table); the others are treated as preconditions in §6.9.

6.2.8 Changes after the inventory

After the day the inventory was made, the product surfaces changed; the catalogue and engine counts in the rest of the chapter (§6.7, §6.8) describe the present state, that is, 22 November 2026. Three changes affect the inventory directly. First, Quarter Focus was removed on 24 October 2026 and the goals module moved to the Life Tree (`goals/lifeTree`) and Sprint (`goals/sprint`) widgets (ADR-035); the `KARAKTER.md` file and rule catalogue of Quarter Focus were deleted, and the files of these two surfaces were written in their place. Second, on 25 October 2026 the Calendar widget (`tasks/calendar`) was added to the Tasks page (ADR-038); the habit check-in list, which the inventory had recorded as a "non-existent surface", was opened with this widget, and the H1–H7 scenarios were transferred there. Third, the count table of the inventory file was updated accordingly: the Quarter Focus row (12 axes, 22 decision points, 22 scenarios) went out; the Life Tree (6, 8, 9), Sprint (6, 11, 8) and Calendar (5, 6, 5) rows came in. The present inventory counts 161 state axes, 313 decision points and 229 scenarios in seventeen files (`Front-end/src/widgets/KARAKTER-IZI.md`, "Sayım (2026-10-24)"); that the scenario total did not change is because 9 + 8 + 5 = 22 scenarios came in to replace the 22 that went out, which is a coincidence. The other fourteen files have not changed since 23 October.

Today there are six more `KARAKTER.md` files under `Front-end/src/widgets/`, but these have no rule catalogue and do not enter the count: Branch Pulse (`goals/pillarPulse`; the component remains, ADR-035), which was removed from the catalogue, Habit Texture (`tasks/habitTexture`), To-do (`tasks/todo`), Projects (`tasks/projects`), Pomodoro (`tasks/pomodoro`) and Circadian Clock (`health/circadianClock`). The file of each of the last four carries a single verdict: the widget reads nothing from the personality profile. The grid catalogue has also grown; `Back-end/config/widgets.php` today defines 26 widgets, including the character analysis card on the overview page (`characterReport`; ADR-073). For this reason the expression "fifteen surfaces" in the chapter belongs to the inventory date, and the expression "seventeen surfaces" to the present rule catalogue.

6.3 The evidence-map review

6.3.1 Design: six strands

The 229 scenarios were distributed across six research strands that are close to one another in subject. The boundaries of the strands were drawn according to the type of difference and the target surface of the scenarios; where a scenario concerned more than one strand (e.g. the sleep plan default is both a chronotype question and a health behaviour question), each strand answered from its own angle and the overlap was resolved in the synthesis. Table 6.4 shows the strands, the questions they cover and the surfaces they feed.

Table 6.4. The six strands of the evidence-map review.

Strand (note file)	Main questions	Surfaces fed
Health behaviour (`kisilik-saglik-davranisi.md`)	`C` and its facets with exercise, medication and weigh-in adherence; low `S` and negative feedback framing; reminder frequency × personality; the effect of tailored interventions; habit duration and streak	Weight, Steps, Water, Workout, Pill Box, Sleep, check-in
Persuasion and message language (`ikna-ve-mesaj-dili.md`)	Personality × Cialdini's principles; message personalisation with the Big Five; gain/loss framing and regulatory focus; praise and celebration; warning tone for low `S`; who deviates from defaults	All language scenarios, celebration and notification
Chronotype and timing (`kronotip-ve-zamanlama.md`)	Chronotype distribution in adults; evening types and exercise; chronotype and eating; sleep hygiene and chronotherapy; social jetlag; notification time and receptivity	Sleep, Workout, Nutrition, Pill Box, Water, Steps, sync
Financial behaviour (`kisilik-finansal-davranis.md`)	`C` with saving and planning; the ostrich effect; envelope budgeting and savings defaults; impulsive spending; spending feedback; financial anxiety; market monitoring and colour	Budget, Vault, Market
Goals, motivation and temperament (`hedef-motivasyon-ve-mizac.md`)	`C` / `S` with goal motivation; progress framing; number of concurrent goals; implementation intentions; streak and fresh start; step size; the evidence status of MBTI/Keirsey, RIASEC and work values	Quarter Focus, check-in, "Pusulan"
Nutrition and weight (`kisilik-beslenme-ve-kilo.md`)	Personality and BMI trajectory; weight loss and maintenance; emotional, uncontrolled and restrained eating; self-weighing; the plateau and the "too fast" warning; communicating the calorie target as a range; body image	Weight, Energy, Nutrition, Body Scan

Note. All notes are under `docs/research/research_notes/Karakter izi kanıt haritası/`; the synthesis is `docs/research/reports/Karakter izi kanıt haritası.md`.

6.3.2 Method and grading

Each strand produced a note with the same structure. The note devotes a section to each of the strand's questions and divides every section into four headings: *Takeaway* (the one-paragraph verdict of the section and its grade), *Cited Findings* (sourced findings, each with a link), *Inferences* (suggestions drawn from the findings for Rehnüma, which the product team needs to test) and *Gaps* (evidence that was sought but not found, or whose numerical breakdown could not be verified). Keeping finding and inference under separate headings is the basic rule of this method; every translation into product defaults was marked as an inference, even where the finding on which it rested was strong.

Access to the sources was mostly at abstract level. Primary abstracts were taken from the Europe PMC, Crossref and Semantic Scholar interfaces; the full text could be read only for studies that were open access or published on an author's page (e.g. Dai et al., 2014; Kivetz et al., 2006; Halko & Kientz, 2010; Bogg & Roberts, 2004). Some publisher sites blocked access; in that case, if the value was taken from a peer-reviewed secondary source it was marked "(secondary)", and if it could not be verified at all it was written under *Gaps*. This honesty rule brought a concrete return: in the persuasion strand, when the primary text of Halko and Kientz (2010) was read, it was seen that the report circulating in secondary sources that "neuroticism is positively related to negative reinforcement" contradicted the primary table, and that the direction was in fact negative (`ikna-ve-mesaj-dili.md`, access note).

The strands used a common four-step scale for the evidence grade. The scale definitions of the strands show small differences according to their own literatures (e.g. in the finance strand large administrative panels and field experiments count as "strong"); the synthesis report reduced these to the common definition in Table 6.5.

Table 6.5. Evidence grade scale.

Grade	Definition	Typical source
strong	Meta-analysis, more than one RCT or a large number of independent replications; direction consistent	Jokela et al. (2013a), Dorison et al. (2022), Jachimowicz et al. (2019)
moderate	A few primary studies, a single large sample or a single RCT; or a replicated perception (self-report) finding	Fischer et al. (2017), Oyibo et al. (2017), Silverman and Barasch (2023)
weak	A single or small study, cross-sectional design, thesis, secondary report; or conflicting findings	Phillips et al. (2021), Webb et al. (2007)
none	Evidence sought but not found	Reminder frequency × C5; temperament × behaviour

Note. A finding may carry two separate grades: "strong" for the main effect, "none" for the interaction of the same variable with language or frequency. The evidence map preserves this distinction in all its tables.

The note to Table 6.5 summarises the most important methodological decision of the chapter. A variable's predicting an outcome (main effect) and a different product behaviour according to that variable giving a better outcome (interaction) were graded separately. For example, the relationship of low `S` with body dissatisfaction is strong (r = .30; Allen & Robson, 2020), but it has not been tested whether the language of self-compassion benefits a user in the low `S` band *more*; the grade is written as "strong (association) / none (language effect)".

6.3.3 Main findings by strand

Table 6.6 summarises the main findings of the six strands together with their effect sizes. The conclusion of each strand that matters for the synthesis is then discussed briefly.

Table 6.6. Main findings and effect sizes of the evidence map by strand.

Strand	Finding	Effect size, sample	Source	Grade
Health behaviour	`C` is related to exercise, but the relationship is small; it is with frequency rather than amount	r = .10 (64 studies); frequency r = .21, amount r = .06	Wilson and Dishman (2015)	strong (small)
	At facet level, activity is predicted most by Industriousness (≈ C4); Self-Control (≈ C5) is weak	r = .18 and r = .07	Bogg and Roberts (2004)	strong
	`C` is related to medication adherence	r = .15	Molloy et al. (2014)	strong (small)
	A notification increases momentary engagement to a small extent; highest at midday at the weekend	RR = 1.04; RR = 1.12	Bidargaddi et al. (2018)	moderate
	Frequent notification does not lower engagement compared with infrequent	d ≈ 0.43–0.50	Morrison et al. (2017)	moderate

Strand	Finding	Effect size, sample	Source	Grade
	No RCT showing the superiority in behaviour of an intervention tailored to personality	—	Alslaity et al. (2023)	none
	Habit automaticity takes long and varies; personality is not a predictor	median 59–66 days, range 4–335 days	Lally et al. (2010); Keller et al. (2021); Singh et al. (2024)	moderate–strong
	Displaying a broken streak lowers subsequent engagement; a streak repair option reduces the harm	66% → 58%, OR = 1.43; with self-attribution OR = 2.80; with repair OR = 2.63	Silverman and Barasch (2023)	moderate
Persuasion and message language	Personality predicts the *perception* of persuasion strategies (`A` → authority, `C` → commitment, low `S` → consensus)	β = .14–.36; R^2 3–15%; N = 216 (Canada), N = 381 (Turkey)	Oyibo et al. (2017); Alkış and Taşkaya Temizel (2015)	moderate (perception), none (behaviour)
	A message matched to personality is perceived as more persuasive; relative increase in clicks in the field	N = 324; 30–50% relative increase in ~3.7 million impressions	Hirsh et al. (2012); Matz et al. (2017)	moderate (direction), weak (magnitude)
	In prevention behaviours gain framing is superior, by a small but consistent margin	r = .083 (94 studies)	Gallagher and Updegraff (2012)	strong
	Loss framing increases anxiety and does not change intention	d = 0.25; N = 15,929, 84 countries	Dorison et al. (2022)	strong
	Self-compassion framing increases the motivation to improve after a mistake	4 experiments	Breines and Chen (2012)	moderate/strong
	In low `S`, a stronger neural response to uncertain feedback than to clearly negative feedback	EEG experiment	Hirsh and Inzlicht (2008)	moderate
	The default effect is large but variable; a "recommended" label and the status quo enlarge the effect	d = 0.68, N = 73,675; b = 0.32 and b = 0.31	Jachimowicz et al. (2019)	strong
Chronotype and timing	Distribution of midpoint of sleep in US adults: early / intermediate / late	40% / 44% / 16%; N = 53,689	Fischer et al. (2017)	moderate
	The performance peak shifts with chronotype	early ≈ 12:00, intermediate ≈ 16:00, late ≈ 20:00; N = 20	Facer-Childs and Brandstaetter (2015)	moderate
	Evening types skip breakfast more often	OR ≈ 1.7	Mazri et al. (2020); Teixeira et al. (2018)	moderate
	Notification response delay is highest at night (00–05)	mean 31 min and 13.6 min; 9.9 million notifications	Stach et al. (2024)	moderate
	Chronotype is stable but shifts earlier with age	r = .56 (7 years)	Druiven et al. (2020)	moderate
	Morningness is related most to `C`; personality explains a small part of the variance in chronotype	ρ = .37, N = 16,647; 11–16% of the variance	Lipnevich et al. (2017)	strong

Strand	Finding	Effect size, sample	Source	Grade
Financial behaviour	`C` is related to active money management and less financial distress; low `S` to financial distress	range r ≈ .13–.37	Donnelly et al. (2012); Xu et al. (2015); Joshanloo (2022)	strong (main effect)
	Not looking at an account that "might be bad" is person-specific and stable; attention peaks on payday	1.2 million accounts; ≈ 30% increase on payday	Sicherman et al. (2016); Olafsson and Pagel (2017)	strong
	High neuroticism increases looking at the portfolio in downturns ("meerkat")	N = 617	Gherzi et al. (2014)	moderate
	Negative relative spending information lowers spending; making access easier reduces fees	12% reduction; 38.4% reduction in fees	D'Acunto et al. (2019); Carlin et al. (2023)	strong
	Impulsive people set a lower budget; showing the remaining budget precisely may increase end-of-period spending	—	Lukas and Howard (2023); Ghosh and Huang (2020)	strong; weak–moderate
	The colour red increases loss focus; frequent evaluation changes risk perception	8 experiments	Bazley et al. (2021); Gneezy and Potters (1997)	strong
Goals, motivation, temperament	The most consistent predictors of goal motivation are neuroticism and `C`	–.31 and +.24; 65 studies; multiple R = .49	Judge and Ilies (2002)	strong
	Monitoring progress increases goal attainment	d = 0.40; 138 RCTs, N = 19,951	Harkin et al. (2016)	strong
	Implementation intentions increase goal attainment; moderation by `C` is conflicting	d = 0.65 (94 tests)	Gollwitzer and Sheeran (2006); Webb et al. (2007); Ajzen et al. (2009)	strong; weak/conflicting
	Temporal landmarks make a fresh start easier	increase of 33.4% on Mondays, 14.4% at the start of the month, 47.1% at the start of term	Dai et al. (2014)	moderate–strong
	The MBTI axes rename four Big Five domains	E–I/E –.74; S–N/O .72; J–P/C –.49; T–F/A .44; N = 468	McCrae and Costa (1989); Furnham et al. (2003)	strong
	Interests predict choice and persistence; overlap with leisure interests is partial	60 studies; N = 3,023, 10 years	Nye et al. (2012); Stoll et al. (2017); Hansen et al. (2008)	strong (career), weak (leisure)
Nutrition and weight	High `C` is protective against obesity	OR = 0.84 per SD; N ≈ 79,000, 9 cohorts	Jokela et al. (2013a)	strong
	Low `S` and low self-discipline are related to emotional eating	N = 442; N = 951	Elfhag and Morey (2008); Keller and Siegrist (2015)	strong
	Flexible restraint is better than rigid	N = 54,517 + 1,838	Westenhoefer et al. (1999)	moderate
	Low `S` is the strongest personality correlate of body dissatisfaction	r = .30; k = 26, N = 39,109	Allen and Robson (2020)	strong

Strand	Finding	Effect size, sample	Source	Grade
	A self-compassion intervention lowers body shame	RCT, N = 228	Albertson et al. (2015)	moderate
	Rapid loss is not regained more than slow loss; dropout is lower with rapid loss	RCT, N = 204; dropout 3% and 18%	Purcell et al. (2014)	strong
	Regular weighing does not on average produce psychological harm; no difference between daily and weekly	r ≈ .00–.02; r+ = .17 in RCTs	Benn et al. (2016); Zheng et al. (2015)	strong; moderate

Note. The effect sizes are in the form reported in the notes; some could be read only from an abstract or a secondary report. Studies whose numerical value could not be verified (e.g. Koo & Fishbach, 2012; Dalton & Spiller, 2012) were taken into the table with the finding on direction, and no value was given.

Health behaviour. The most important conclusion of this strand is negative: no RCT could be found showing that a digital health intervention tailored to personality outperforms an untailored one in the behavioural outcome (Alslaity et al., 2023). The evidence available is at the level of preference and perception and is cross-sectional. The second conclusion directly undermines the first example of ADR-025 §7. At facet level, activity is predicted most by industriousness (≈ C4); self-control (≈ C5) is a weak predictor and is related rather to avoidance behaviours such as alcohol, substance use and risky driving (Bogg & Roberts, 2004). The only direct study, in which reminder duration did not differ by Conscientiousness, points the same way (Dennis et al., 2015). The hypothesis "frequent reminders in low C5" therefore has no direct support. The third conclusion is that `C` has more to do with "how regularly" than with "how much" (frequency r = .21, amount r = .06; Wilson & Dishman, 2015); for low `C`, the product problem should be framed as continuity and not intensity. Finally, the individual range of the time to automaticity (4–335 days) and the fact that only 23% of participants reached the habit threshold (Singh et al., 2024) show that widgets should not promise a number such as "~66 days"; in the single cross-sectional study, `C` was found, contrary to expectation, to be negatively related to perceived automaticity.

Persuasion and message language. The relationships between personality and the perception of persuasion strategies are small to moderate in size and have been partly replicated in two countries: the Canadian sample replicated 9 of the 18 Turkish findings (Oyibo et al., 2017). The recurring pattern is that high `A` and low `O` are more open to social strategies such as authority, consensus and liking, whereas high `O` and low `S` respond positively to almost none of the common strategies (Ndulue et al., 2022). These are perceptions; they have not been tested at the level of behaviour. The conclusions of this strand that are really fruitful for the product are independent of personality: that loss framing increases anxiety and does not change intention (Dorison et al., 2022), that self-compassion framing increases the motivation to improve (Breines & Chen, 2012), that uncertain feedback produces more reaction than clearly negative feedback (Hirsh & Inzlicht, 2008) and that inflated praise increases the avoidance of challenge in children with low self-esteem (Brummelman et al., 2014). On defaults, the only direct personality evidence runs contrary to expectation: in a large German sample, the preference for the preselected option increases with high `O` (Grüner & Rohde, 2022). That the negative dimension of self-compassion is almost identical with neuroticism (r ≥ .85) suggests that the low `S` user is the person most prone to self-criticism; this is an inference and does not take the place of an interaction test.

Chronotype and timing. Rhythm is the only variable with a moderate grade of evidence for shifting time defaults. The evidence is in two parts. The first is that chronotypes really do differ in their sleep times and performance peaks (Fischer et al., 2017; Facer-Childs & Brandstaetter, 2015). The second is missing: no study could be found that directly tests the questions "does a default plan adjusted to chronotype increase adherence" or "does chronotype moderate receptivity to notifications". The 2026 RCT of chronotype-aligned exercise (N = 150) reported greater gains in blood pressure and sleep quality, but excluded intermediate types (Open Heart, 2026). The measurement caveat of the strand is also important: most of the literature works with the continuous midpoint-of-sleep measure of the MCTQ, whereas the product's three-class short scale is close to the MEQ tradition, and the Turkish correlation between the two measures is only r = −.47. The conversion from class to clock time is therefore coarse. The common principle of the strand is that times are derived relative to the user's sleep plan and not from the chronotype class (wake-up + N, bedtime − N), and that Rhythm shifts only the initial default of the sleep plan.

Financial behaviour. `C` turns out to be closer to process (management, not experiencing distress), and `E` and `O` to the balance sheet (assets, debt) (Brown & Taylor, 2014). The ostrich effect has been replicated in three large data sets and found to be person-specific and stable over time (Sicherman et al., 2016). The only study that measures a personality moderator shows that low `S` does not automatically mean "does not look": high neuroticism can increase looking in downturns (Gherzi et al., 2014). This finding refines the direction of the ADR-025 §7 example, "gentle language against the ostrich effect in low Emotional stability": the problem may be not only avoidance but, depending on the context, avoidance or obsessive looking. The field evidence points one way: negative relative information changes behaviour, and hiding information does not help. Loss and gain framing make no difference in reminders; by contrast, a monthly reminder that names the goal increases savings by about 6% (Karlan et al., 2016). There is no study at all on personality × envelope ratio preference; the finding that impulsive people already set a lower budget (Lukas & Howard, 2023) is a reason for protecting, not lowering, the savings share in low S5.

Goals, motivation and temperament. This strand did two separate jobs. First, it confirmed that `S` and `C` are the most robust personality predictors of goal motivation (Judge & Ilies, 2002) and compiled personality-independent mechanisms such as progress monitoring, framing, implementation intention and the fresh start. In none of these mechanisms has moderation by personality been directly shown; in the implementation intention it is conflicting (in Webb et al., 2007, the effect is larger in low/medium `C`; in Ajzen et al., 2009, high `C` carries out its intention more). The only consistent boundary condition is that plans do not work under high impulsivity (urgency) (Churchill & Jessop, 2011). Second, the strand examined the evidence status of Pusula's derived layers; these conclusions are treated separately in §6.4.

Nutrition and weight. The weight of the evidence gathers on two axes: `C` (especially C5 and C2) is the most consistent predictor of obesity and weight gain (Jokela et al., 2013a); low `S`, for its part, is related less to the prevalence of obesity than to weight fluctuation, emotional eating and body dissatisfaction (Elfhag & Morey, 2008; Allen & Robson, 2020). Two findings of the strand affect the product copy directly. The first is that the plateau in the sixth month is explained by intermittent non-adherence rather than by metabolic adaptation (Thomas et al., 2014). The second is that rapid loss is not regained more than slow loss and that dropout is lower in the rapid group (Purcell et al., 2014). This second finding shows that the ">1%/week" warning cannot be defended on the ground of "regain"; the warning can be given only on the ground of nutritional adequacy, and only once. The note also

carries a methodological caveat: most of the literature assumes calorie or meal *logging*, whereas on the nutrition side Rehnüma shows a plan and keeps no consumption log (ADR-013). For this reason, findings that assume logging were not carried over directly to the plan surface.

6.3.4 Synthesis: resolving contradictions between the notes

Because the six notes were written independently, at some points they produced suggestions that run counter to one another. The synthesis report put these explicitly on record and applied a resolution rule: neither of two conflicting inferences produces a rule, and both become measurement items. The clearest example is reminder frequency. While the chronotype note proposed the principle "the hour from Rhythm, the frequency from C5", the health and persuasion notes found no frequency evidence for C5 or `C`. The synthesis ruled that the principle is right for *separation* (the two variables should not be multiplied with each other) but insufficient for a frequency rule; reminder frequency × C5 became an A/B candidate. The other unresolved contradictions were handled in the same way: the negative relationship between weighing and self-esteem (Pacanowski et al., 2015) and the meta-analyses that found no harm (Zheng et al., 2015; Benn et al., 2016); the small study in which neuroticism was found to be related to *more* loss on a very-low-energy diet (Munro et al., 2011); extraverts' liking for the leaderboard (Jia et al., 2016) and their being demotivated by reward and competition (Ndulue et al., 2022); the `C`–activity relationship being strong in the young (Bogg & Roberts, 2004) and the protective effect increasing with age (Sutin et al., 2016). None of them produced a rule.

6.4 Verdict by variable

6.4.1 Three permitted uses

The evidence map gave every Pusula variable one of three permitted uses: **K** (a rule may be built), **T** (tone or A/B candidate only) and – (not used, narrative only). The permission was given by looking together at the evidence grade of the variable, its report tier and the direction of the proposed effect. Table 6.7 summarises the verdict.

Table 6.7. Verdict on use for the Pusula variables.

Variable	Tier	Permission	For what	Basis	Grade
`S` band	summary	K	Softening the language of warning, overspend and "behind"; putting the trend forward; removing exaggeration from celebration	Relationship with emotional eating, body dissatisfaction, financial distress and motivation (Elfhag & Morey, 2008; Allen & Robson, 2020; Xu et al., 2015; Judge & Ilies, 2002)	strong (main effect), none (interaction)
S1 Managing worry	full	K	"Number + a single way out", leaving no uncertainty; proxy for financial anxiety	Financial anxiety and information avoidance (Shapiro & Burchell, 2012); competence-preserving feedback (Kappes et al., 2012)	moderate
S3 Mood balance	full	K	Plainness of celebration; not emphasising setbacks	Inflated praise backfiring in low self-esteem (Brummelman et al., 2014)	moderate (the proxy mapping is an assumption)

Variable	Tier	Permission	For what	Basis	Grade
S5 Impulse control	full	T	Normalising overspend in Wants; trigger-moment reminder	Urgency and compulsive buying (Billieux et al., 2008); boundary condition of plans (Churchill & Jessop, 2011)	moderate (buying), weak/conflicting (dropout)
`C` band	summary	K	Default for logging friction and information density; risk flag	Obesity OR = 0.84/SD (Jokela et al., 2013a); medication adherence r = .15 (Molloy et al., 2014)	strong (main effect, small)
C5 Self-discipline	full	T	Not used as a condition for reminder *frequency*; only an "increase structure" candidate	Self-Control–activity r = .07 (Bogg & Roberts, 2004); reminder duration no different by `C` (Dennis et al., 2015)	weak, none (frequency)
C2, C4, C6	full	T	Note field, visibility of numerical progress, tolerance band	Indirect (Gold et al., 2020; Cianci et al., 2010); no evidence for C6	weak–none
`E` / E6 Cheerfulness	summary / full	T	Liveliness of celebration (tolerance only, not an increase)	Conflicting (Jia et al., 2016; Ndulue et al., 2022); wearable adherence OR = 0.67 (Martínez et al., 2021)	weak/conflicting
`A` band	summary	T	Tone of expert reference and social proof	Two countries at the level of perception (Oyibo et al., 2017; Alkış & Taşkaya Temizel, 2015)	moderate (perception), none (behaviour)
`O` / O5 Curiosity about ideas	summary / full	T	Length of rationale; not showing social comparison	Negative β (Oyibo et al., 2017; Ndulue et al., 2022); need for cognition as a framing moderator (Covey, 2014)	moderate (perception)
Rhythm	full	K	Initial default of the sleep plan (±60 min); times derived from the plan	Fischer et al. (2017); Facer-Childs and Brandstaetter (2015); Druiven et al. (2020)	moderate (shifting), none (increase in adherence)
Interests (RIASEC first letter)	summary	T	*Order* of the empty-state and first-goal examples	Nye et al. (2012); Hansen et al. (2008)	strong (career), weak (leisure)
Mind motif (4 axes, 16 codes)	full	–	No rule	The axes are a renaming of the domains (McCrae & Costa, 1989); type unstable (Pittenger, 2005)	none (incremental value)
Temperament SP/SJ/NF/NT	full	–	No rule	No study of Keirsey × Big Five or behaviour	none
Self type (8 archetypes)	summary	–	Only the "Pusulan" heading	Derived label; fragile at retest (Chapter 4, Gate 0)	none
Natural tendencies (10)	full	–	Narrative	Combination of facets; no separate evidence	none
Work values (6)	full	–	Narrative	Person–environment fit limited to job satisfaction (Kristof-Brown et al., 2005)	none
Quality	summary	gate	`uninterpretable` → no rule; `caution` → personality-dependent rules off	The harm of a mismatched message has not been tested (Sharp et al., 2018)	—

Note. In adapted and abridged form: `docs/research/reports/Karakter izi kanıt haritası.md` §2. The Tier column shows whether the variable arrives in the summary or in the full report (ADR-025 §6). For S2, S4, S6, C1, C3, A1, A5, A6, E4, E5 and O2–O4 the permission is T and the grade is weak or none; no separate row is given in the table. The table is

the verdict of the evidence map as of 23 October 2026; the addendum to ADR-033 dated 24 October 2026 permitted rules of the "suggestion copy" type for the interest first letter, Rhythm and the `C` band (§6.6.1).

The overall picture of Table 6.7 is clear when it is compared with the examples in ADR-025 §7. Of the four examples only two survive at the level of a rule: gentle language in low `S` (not limited to the budget, but on all warning surfaces and only in the direction of attenuation) and time according to chronotype (relative to the plan, not from the class). "Frequent reminders in low Conscientiousness" has no evidence behind it; "a smaller KR step", for its part, ceases to depend on personality and turns into a principle applied to everyone (the first step is small and can be taken at once; Kivetz et al., 2006). In the present catalogue the only personality-dependent trace of this example is a piece of suggestion copy that appears in the Life Tree builder in the low `C` band (`lifeTree.S3.lowC-small`); the copy changes no value (§6.6.1). No evidence could be found for "workout variety in high Openness".

6.4.2 Why derived layers cannot be conditions

The layers of the Pusula report that appeal most to the reader (mind motif, temperament, archetype, natural tendency, work value) are those with the least evidence for the rule engine. Why these layers cannot be the condition of any widget rule rests on four grounds.

Lack of incremental validity. The four motif axes are a renaming of four Big Five domains. McCrae and Costa (1989) report correlations between the MBTI indices and the NEO domains of −.74 for E–I ↔ `E`, .72 for S–N ↔ `O`, −.49 for J–P ↔ `C` and .44 for T–F ↔ `A`; Furnham et al. (2003) replicate this in a sample of N = 900. Given that Pusula derives the motif from Big Five scores in the first place, a rule written on the motif would at best be a coarser copy of a rule that could be written on the domain band. That the relationship of the T/F axis with `A` is only .44 shows that this axis carries both its own theoretical meaning and its Big Five counterpart weakly.

Categorical instability. The score distributions are unimodal; there is no support for a "true dichotomy", and the type changes in 39–76% of cases at a five-week retest (Pittenger, 2005). Stein and Swan (2019) conclude that the theory does not agree with known data, is untestable and is internally contradictory. The Gate 0 simulation in Chapter 4 had shown a similar fragility for Pusula's own derived labels as well. A rule tied to a categorical label changes the plan of a user close to the boundary through measurement error.

Lack of behavioural prediction. For the Keirsey temperament not even a study of its correlation with the Big Five could be found; the KTS-II overlaps only with the MBTI (Kelly & Jugovic, 2001). There is no study showing that temperament predicts any everyday behaviour beyond the domains (`hedef-motivasyon-ve-mizac.md` §7). The archetypes and natural tendencies are combinations of facets; they have no separate evidence base.

Lack of transfer across contexts. The evidence for work values is at the level of the person–environment fit meta-analysis and is limited to job satisfaction (Kristof-Brown et al., 2005); there is no direct study of its transfer to everyday goal motivation or to budget preference. In the source scale the test–retest reliability of the work values is in the range r = .59–.66. The situation is better for RIASEC, but limited: interests strongly predict choice and persistence in the career context (Nye et al., 2012; Rounds & Su, 2014), whereas their overlap with leisure interests is partial (Hansen et al., 2008). For this reason the interest code can change only the *order* of the empty-state examples, cannot restrict the content, and can do even that only as an A/B candidate. This is the verdict of the evidence map.

The addendum to ADR-033 dated 24 October 2026 later accepted with `rule` status six records that show an invitation text according to the interest first letter in the empty career branch of the Life Tree (§6.6.1); these records do not restrict the content either, and only add text to an empty area.

These four grounds were turned into a product decision in ADR-033 and made executable in three layers: the signal class of the rule engine does not produce these fields at all (`Backend/app/Services/PersonalityRules/PersonalitySignals.php`), the catalogue validator rejects the conditions `motif`, `temperament`, `archetype`, `tendency` and `value` (`RuleCatalog::FORBIDDEN_VARS`), and `InvariantsTest` tests both that the validator catches the prohibition and that no record in the catalogue depends on these variables (§6.7.5). The decision records temperament-based personalisation ("meaning for the NF, mechanism for the NT") as a rejected alternative and notes that it would conflict with the trust design in ADR-031. The scope of the rejection is narrow: these layers remain in the report and in the "Pusulan" narrative; they merely cannot be the condition of widget behaviour. In the inventory matrices these columns were not deleted but filled in on every row as "narrative only"; in this way the reader can see in every scenario why temperament does not change anything.

The list of derived layers that cannot be conditions grew longer with form v5. Form v5 adds two new layers to the report: the self-view (`self`), which comes from the user's self-placement among the eight archetype cards, and the expression profile (`expression`), which comes from the statement ranking; the report also gives their agreement with the measured profile (`agreement`). The sixth clause of ADR-044 decides that the rule engine does not change, and that the self-view and the expression profile cannot be rule conditions and are narrative only. The rationale is the same: both are categorical and derived layers, and there is no evidence that they predict widget behaviour. In the code `RuleCatalog::FORBIDDEN_VARS` today carries ten root names: `motif`, `temperament`, `archetype`, `tendency`, `value`, `values`, `code`, `self`, `expression` and `agreement`. `PersonalitySignals` produces none of these fields; the ADR-044 test in `InvariantsTest` separately tests that the last three names are in the forbidden list.

6.5 Robust principles independent of personality

The most productive output of the evidence map is a set of principles that can be applied to every user without looking at personality at all. They take precedence over personality-dependent rules for two reasons. First, their evidence is stronger: most rest on a meta-analysis, a field experiment or a large RCT. Second, they are robust to measurement error: even if the user's profile has been mislabelled, the rule does not change. For this reason ADR-033 explicitly rejected the option of "leaving everything to A/B testing"; waiting on strong rules such as removing loss language or not emphasising a broken streak is, in the words of the evidence map, "harm". Table 6.8 lists these principles with their sources and grades.

Table 6.8. Robust principles independent of personality.

Principle	Product counterpart	Evidence	Source	Grade
Framing rule	Between 0% and 50% progress the "covered" part leads; after 50% the "remaining" part leads	Small-area hypothesis	Koo and Fishbach (2012)	moderate

Principle	Product counterpart	Evidence	Source	Grade
Loss language is not recommended	Gain framing in warnings; no "you will lose" language	Gain framing is slightly but consistently superior in prevention; loss framing raises anxiety and does not change intention; no difference in reminders	Gallagher and Updegraff (2012); Dorison et al. (2022); Karlan et al. (2016)	strong
Self-compassion tone	Mistakes and setbacks are described as common, human and changeable	Four experiments; RCT on body shame	Breines and Chen (2012); Albertson et al. (2015)	moderate/strong
Leaving no uncertainty	Not "it may be risky" but a number and a duration; number + a single way out	Larger response to uncertain feedback; negative information changes behaviour; competence must be protected	Hirsh and Inzlicht (2008); D'Acunto et al. (2019); Kappes et al. (2012)	moderate
"You are doing well" gives licence	The remaining target instead of praise ("X ₺ left to your savings goal")	Perceived progress frees the person to turn to the conflicting goal	Fishbach and Dhar (2005)	strong
A plateau is normal	A plateau is "an expected period, the sum of small deviations"; the word "non-adherence" is not used; the ">1%/week" note appears once, justified by nutritional adequacy	The plateau is explained by intermittent non-adherence; rapid loss is not regained more	Thomas et al. (2014); Purcell et al. (2014)	moderate; strong
No habit duration is promised	"~66 days" is not shown; "missing a single day does not break the process" may be used	Median 59–66 days, range 4–335; 23% reach the threshold	Lally et al. (2010); Keller et al. (2021); Singh et al. (2024)	moderate–strong
A broken streak is not emphasised	When the streak breaks, the counter is not emphasised; streak repair and a fresh start are offered	A broken streak lowers subsequent engagement; repair reduces the harm	Silverman and Barasch (2023); Lally et al. (2010)	moderate
Colour and frequent evaluation	A fall is shown in amber/neutral + arrow; saturated red only at a large threshold; daily % on the second line, 30-day direction first	Red increases loss focus; frequent evaluation changes risk perception	Bazley et al. (2021); Kliger and Gilad (2012); Gneezy and Potters (1997)	strong
Presentation of defaults	The default comes with a "recommended" label and a one-sentence rationale, and changes with a single tap	The "recommended" label and the status quo enlarge the effect; short decision time strengthens the default	Jachimowicz et al. (2019); White et al. (2021)	strong
Fresh start	A "start again on Monday" suggestion for the user whose streak has broken or who has dropped the programme	Goal behaviour increases on Mondays and at the start of the month and of the term	Dai et al. (2014)	moderate–strong
Payday	Period summary and reflection on payday; when there is no income, an "estimated" label and a countdown to payday	Financial attention rises by ≈ 30% on payday	Olafsson and Pagel (2017)	strong
Night silence	No notifications are sent between 00 and 05; times are relative to the plan	Response latency is highest at night	Stach et al. (2024); Bidargaddi et al. (2018)	moderate

Note. Source: `docs/research/reports/Karakter izi kanıt haritası.md` §3. Each principle is represented in the catalogue by at least one record with an empty `when.all` and `status = rule`.

6.5.1 Language principles

The first four principles set the product's voice and generate most of the unconditional rules in the catalogue. The framing rule changes the sentence of progress indicators according to the stage; in the measurement form of Body Scan, the counter leads with the measurements entered in the first half of the form (the "5 of 8 measurements entered" pattern) and with the measurements remaining in the second half (`bodyScan.S1.progress-frame`). The ban on loss language turns the sentence "You are 600 ml behind" in the water widget into "600 ml left · 5 hours to go. One glass will make it." (`waterTracker.S4.remaining-frame`). The two sentences give the same information; the second adds a time context and an attainable action to it, and does not emphasise the loss.

The principle of leaving no uncertainty works together with the "number + a single way out" pattern. For a budget overrun the unconditional rule uses this pattern: "{over} ₺ over plan · {daysLeft} days left. Spending {perDay} ₺ a day less on wants evens it out by the end of the period." (`budget.S1.number-and-exit`). The first part of the sentence reports the situation with a number; the second part gives a single, calculated way out. This pattern is the evidence-map counterpart of the "yellow tone + way out" decision in ADR-024 §5. In Energy Needs the same principle turns formula uncertainty into a range stated without modal hedging: "Formulas give an average; your real expenditure is most likely between 1,980 and 2,420 kcal." (`energyNeeds.S7.range-in-numbers`).

The self-compassion tone is used most on the body and setback surfaces. In the budget, the unconditional copy for a period in which no share could be set aside for savings reads "No share went to savings this period; this happens often and can change. Setting a small amount aside on payday next period would be the first step." (`budget.S9.self-compassion`). The sentence carries the self-compassion components of Breines and Chen (2012) (common humanity, changeability) and ties the payday principle to the suggested action.

6.5.2 Visual and ordering principles

The broken-streak principle is the only point at which the evidence map comes into direct tension with an existing ADR. ADR-004 had decided that there is no "grace day" in the habit streak. The seven-study finding of Silverman and Barasch (2023) shows that displaying a broken streak lowers continuation despite identical past behaviour (from 66% to 58%) and that a repair option reduces this harm. ADR-033 recorded this tension as open work: before the streak surface is opened, ADR-004 must be updated with a streak repair or fresh start option. This update was made on 25 October 2026 (addendum to ADR-004, ADR-038): there is still no tolerance day, but a check-in made after the fact for yesterday or the day before counts that day as completed and the streak is calculated accordingly ("streak repair"). The Calendar widget opened by the same decision leaves a missed day empty and shows no broken-streak warning (`tasks/calendar/KARAKTER.md`). In the catalogue the principle is represented by unconditional rules: in Steps and Water Tracking a broken streak is shown neutrally and a fresh start is offered (`stepCounter.S10.broken-streak-quiet`, `waterTracker.S11.broken-streak-quiet`). These records combine three principles: not emphasising the break (Silverman & Barasch, 2023), saying that a single miss does not break the process (Lally et al., 2010) and suggesting a fresh start window (Dai et al., 2014).

The colour principle changes the default appearance of the finance surfaces. In Vault the loss line is written in an amber or neutral tone in the form "▼ {amount} ₺ from purchase · {pct}%" (`vault.S3.neutral-wording`); saturated red is used only when a large threshold is exceeded. The find-

ing that red increases loss focus and that this effect is not observed in colour-blind people (Bazley et al., 2021) shows that the effect stems from the colour itself; the red = fall code is kept, and only a saturation threshold is added.

6.5.3 Default and timing principles

The principle on the presentation of defaults does not suggest different values by personality; it regulates how the *same* value is presented. In budget setup the 50/30/20 split is the same for everyone and arrives with the sentence "Recommended: 50/30/20. A fifth of your income set aside, the rest for needs and wants. You can change it any time." (`budget.S5.recommended-label`). The same pattern is repeated in the water goal ("Recommended: by your weight, 35 ml per kg."), the step goal ("The recommended start is 8,000 steps a day") and the energy pace ("Moderate · recommended."). The meta-analysis of Jachimowicz et al. (2019) shows that the "recommended" label and status quo framing enlarge the default effect, while ease of change does not reduce it; user autonomy is preserved without the guiding power of the default being lost.

For the timing principles the common rule is that times are derived relative to the sleep plan, together with night silence. These principles do not depend on personality: Rhythm changes only the *initial* value of the plan (§6.6), and the calculation of times derived from the plan is the same for everyone. For example, as an unconditional rule the workout suggestion says "Suggestion: 16:00–18:00. Your body is usually more ready at these hours; you can pick the time you want." (`workoutPlan.S7.suggestion-copy`); the sentence makes no adherence claim, because it has not been shown that a chronotype-aligned time increases *adherence*.

6.6 Evidence policy

The evidence map is a compilation of findings; for it to become product behaviour, a policy is needed that determines which finding becomes a rule under which condition. We wrote the policy in the fifth section of the synthesis report and made it binding through ADR-033. The policy consists of five parts: status thresholds, quality and tier gates, the neck-and-neck band, the conflict order, and transparency and override; a measurement plan accompanies them.

6.6.1 Three statuses

Each record receives one of three statuses. Table 6.9 shows the conditions of the statuses.

Table 6.9. Rule statuses and assignment conditions.

Situation	Status	Condition
Personality-independent evidence is strong or moderate	`rule` (unconditional)	`when.all` empty; for every user, whatever the quality
Personality-dependent; main effect strong **and** the effect is attenuation/softening only (hides no information, changes no value)	`rule` (conditional)	Quality `ok`, variable available at the tier, can be undone with one click
Personality-dependent suggestion copy: the default is empty and the rule only adds text; evidence at least moderate (ADR-033 addendum)	`rule` (conditional)	Quality `ok`; changes no value, removes no option, hides no information; the text refers to the user's measured characteristic and contains no persuasion technique; a "Character trace" label and a one-sentence rationale in the interface

Situation	Status	Condition
Personality-dependent; main effect moderate/weak or entirely inference; or the effect is an increase or a change of value	`ab_candidate`	Born closed; opened only with a defined measurement plan and an ADR
No evidence; forbidden variable (motif, temperament, archetype, tendency, work value); or conflicting evidence	`no_change`	"No change" record, with its rationale

Note. Source: `docs/research/reports/Karakter izi kanıt haritası.md` §5.1; ADR-033 §2 and its addendum of 24 October 2026 (third row). Numerical thresholds ("how many goals", "what % step", "how many minutes before") appear in no source; in the catalogue such values are marked as "default without evidence" in the `evidence.note` field.

The second row of the table is the finest point of the policy. For a personality-dependent rule to receive `rule` status, two conditions must hold together: the main effect must be strong and the effect must be in the direction of attenuation only. Attenuation was defined by three properties: the rule hides no information (the number always remains visible one tap away), changes no value (the goal, the ratio, the time stay the same; only the tone, the order or the visual emphasis changes) and can be undone with one click. The rationale for this definition is asymmetric risk. When a softening rule is applied to a mislabelled user the cost is small: a balanced user reads a somewhat calmer sentence. When a value-changing rule is applied to a mislabelled user, by contrast, the user receives a wrong plan (e.g. a lower savings ratio). Given that labels can change at a rate of 43–55% at retest (ADR-029, Gate 0), the second kind of rule cannot be opened without measurement (ADR-033, Rationale).

A direct consequence of this policy is also written in the verdict of the evidence map: no scenario that suggests a *different value* by personality (envelope ratio, goal pace, number of reminders, sleep duration chip, number of goals) gains rule status; all of them are A/B candidates or "no change" records.

The third row of the table is not in the evidence map; it was added to ADR-033 while the Life Tree was being designed (24 October 2026, ADR-035). The addendum extends the definition of `rule` to a single case outside attenuation: suggestion copy added to an empty area. A personality-dependent record may be a `rule` if it satisfies all five conditions: the default is empty and the rule only adds text; it changes no value, removes no option, hides no information, and the user can pass it by without reading; the text refers to the user's own measured characteristic (interest code, Rhythm, domain band) and contains no persuasion technique such as scarcity or social proof; the evidence grade is at least moderate and its source is in the evidence map; it arrives in the interface with a "Character trace" label and the rationale sentence is present in the record. The rationale for the addendum is that these records make no claim of a behavioural effect; suggestions with weak evidence still remain `ab_candidate`. In today's catalogue, records of this kind are in the Life Tree and Calendar: six invitation texts by the first letter of the interest code on the empty career branch, three invitation texts by Rhythm on the empty health branch, two time examples by Rhythm in the habit composer and two "start small" suggestions in the low `C` band (§6.7.2). This extension also loosens one limit of the verdict in §6.4: the first interest letter and the `C` band now appear in conditional `rule` records. In terms of evidence these are records that rest on a main effect and whose interaction has not been tested; the limitation in §6.9 applies to them as well.

6.6.2 Quality and tier gates

Quality gate. If the profile quality is `uninterpretable`, no rule carrying a personality, Rhythm or interest condition is evaluated; the widget keeps its baseline state. If the quality is `caution`, personality-dependent `rule` and `ab_candidate` records are closed and only unconditional rules run. The evidence map ties this decision to two reasons: the harm of a message mismatched to personality has not been measured (Sharp et al., 2018), and derived labels are fragile at retest. Pusula-API reports quality at four levels (`ok`, `note`, `review`, `uninterpretable`); the engine maps these onto three gates: `ok` and a single-flag `note` count as `ok` for the engine, whereas a `review` with two or more flags counts as `caution` (ADR-033 §3; `PersonalitySignals::qualityLevel`). The engine applies this gate centrally, not at record level: a conditional rule is evaluated only while quality is `ok`; the catalogue author does not need to write this separately into each record (`Back-end/database/data/personality_rules/README.md`).

Tier gate. The summary (stage 1) profile contains only domain bands, interests and quality; facets and Rhythm require stage 2 (ADR-025 §6). Because the result is no longer shown without an account (ADR-073), the tier expresses not whether the user is registered but which stage of the test has been completed; the engine looks only at the `stage` field of the profile. For this reason only domain-band rules (e.g. the `domain.S = low` softenings) and interest ordering can run at the summary tier. If the variable of a condition is not present in the profile, the condition counts as "not satisfied"; for example, a rule conditional on `facet.S1` is never triggered on a stage 1 profile. Even in the full report the motif, temperament and value fields are not read by the engine; this is not a tier rule but a schema decision.

6.6.3 The neck-and-neck band: 40–60

If a facet percentile lies between 40 and 60, the variable counts as `mid` (neck-and-neck) and triggers no rule conditional on `low` or `high`. The thresholds are kept in `Back-end/config/personality_rules.php` as `low_below = 40` and `high_above = 60`. If there is no percentile, Pusula's own band is used. This fallback path was written for the case in which, while percentiles were computed from Rehnüma's own reference group, the group remained below 30 people (ADR-032; ADR-033 §3). Because percentiles have come from the published norm of the IPIP-NEO-120 since 22 November 2026 (ADR-082), in profiles scored with this norm the domain and facet percentiles are available regardless of sample size; the exception is the O6 facet, which has no counterpart in the norm data, and in these profiles the fallback path runs for it alone. In old-format records (before form v4) the band is derived from the 0–100 domain score with the same thresholds (`PersonalitySignals`).

The width of the band stands in for a confidence interval. Facet scores are computed from four items and their standard errors of measurement are larger than those of domain scores (Chapter 3, Chapter 4). To count a user as `low` or `high` at a percentile close to the middle would be to turn a difference lying within measurement error into product behaviour. The twenty-percentile-point neck-and-neck band ensures that the rule engine comes into play only in the tails of the distribution, that is, in the regions where the signal is relatively less affected by measurement error. The band is also consistent with the principle of report honesty in Chapter 5: a result presented as "neck-and-neck" in the report is not used as if it were a tendency on the dashboard.

6.6.4 Conflict order

More than one source may suggest a value for the same decision point: the user's own setting, measured behaviour, an unconditional rule and one or more personality rules. The policy places these in a fixed order of precedence (ADR-033 §4):

1. **The user's setting overrides the rule.** If the user has changed a default or has pinned a surface with "keep it this way", no rule is applied for that key; the pinning is recorded with the `rule_overridden` event.
2. **Measured behaviour overrides Rhythm.** If there are at least 14 consistent nights of sleep data, the average bedtime (`avg_bedtime`) determines the plan default and the times derived from it. The rationale is that chronotype is only moderately stable over seven years ($r = .56$) and shifts with age (Druiven et al., 2020). The same logic holds in finance: if a user-specific "looking" pattern has been recorded over several periods, this is a more direct signal than the `S` band.
3. **An unconditional rule overrides a personality rule.** A personality rule cannot produce the opposite of an unconditional rule; loss language, broken-streak emphasis or information hiding is not opened for any band.
4. **Two rules on the same key.** If a conditional and an unconditional rule target the same `effect.key`, the conditional one (the more specific one) wins; this means that the *softened version* of the unconditional rule comes into play, and it does not contradict item 3, because a conditional rule can in any case only attenuate. If two personality rules are on the same key, the one in the direction of attenuation wins.
5. **No products of variables are written.** Combined conditions such as `S` × Rhythm are forbidden; each variable is read as a separate rule. That personality explains only 11–16% of the variance in chronotype (Lipnevich et al., 2017) shows that the two variables should not be substituted for each other but read separately. If age is available, the Rhythm shift is reduced with age; no age × `C` rule is written, because the relevant findings conflict.

The first two items of this order deliberately push personality information into the background. The profile is positioned as the *first guess* made when nothing is known about the user; when the user says something or behaviour is measured, the guess withdraws.

6.6.5 Transparency and override

Each rule is keyed by an identifier (`id`) and each rule has a short rationale sentence (`copy.why`); the sentence ends with the source key, for example "A single-weigh-in difference is mostly water and salt fluctuation; the trend difference produces less anxiety (Benn2016)." (`weightTrend.S2.toast-trend-delta`). The policy envisages that on every surface where a rule is applied, a "Why are we showing it this way?" link shows this sentence and two buttons: "Back to default" and "Pin this surface". In this way every personality-dependent behaviour becomes visible to the user, justified and reversible. A rule with `ab_candidate` status is born closed and is opened only by an ADR. A schedule was also proposed for asking about Rhythm again: once a year under the age of 30, once every two to three years above it, or an update as sync data arrives.

Applying this link and these buttons to the widgets is outside the scope of ADR-033; the decision leaves the application of rules to surfaces as separate pieces of work per widget (§6.7.4, §6.9).

6.6.6 Measurement plan

For A/B candidates to be opened, and for it to be seen whether personality-dependent rules actually work, a series of events that are not recorded today must enter the schema. Table 6.10 summarises the measurement plan.

Table 6.10. Character trace measurement plan: events and the hypotheses they test.

Event	Fields	Hypothesis tested
`widget_view`	Opening time, state at that moment (`behind`, `over`, `plateau`), time elapsed since the previous bad state	Person-specific ostrich coefficient (Sicherman et al., 2016); avoidance or obsessive looking at low `S`
`widget_dismiss`, `surface_pinned`	Surface, rule active at that moment	Rule rejection rate; the "softening is unnecessary" signal
`reminder_sent`, `reminder_opened`, `reminder_acted`, `reminder_snoozed`	Plan-relative sending time, latency, time class	Frequency × C5 (A/B); time × Rhythm; night latency
`default_changed`	Key, old and new value, time since setup	Hypotheses on defaults by personality; the effect of the "recommended" label
`celebration_shown`, `celebration_skipped`	Intensity variant, action taken afterwards	Plain versus lively celebration × S3/E6
`copy_variant_shown`	Variant, action within the following 24 hours	Language variant × `S` / `A` / `O`
Weigh-in interval, check-in streak, plan adherence	Sleep ±30 min, meal time	Behavioural outcome; monitoring effect (Harkin et al., 2016)

Note. Each event carries the `widget`, `scenario`, `rule_id` and active band fields. Source: `docs/research/reports/Karakter izi kanıt haritası.md` §5.6.

The primary criteria are the **first-opening latency** of the widget after a bad state (overrun, falling behind, plateau) and the **seven-day action rate**; the secondary criteria are the rule rejection rate and the default change rate. In the product's present single-user (dogfooding) context a classical A/B test is not possible; the plan instead envisages an ABAB comparison across periods: the rule on for one period, off for one period, then on and off again. No `ab_candidate` is promoted to `rule` status without at least two periods of data. A single-person ABAB design has no external validity; it says only whether the rule works for that user. This limitation is discussed separately in §6.9.

6.7 Implementation

6.7.1 Rule schema

The policy is held as data in a catalogue consisting of one JSON file per surface (`Backend/database/data/personality_rules/<widget>.json`). The rationale for separating the catalogue from the engine is written in ADR-033: as the evidence changes the record changes, the engine does not; each record carries its own source. Table 6.11 shows the fields of a record.

Table 6.11. Fields of a rule record.

Field	Content	Contract
`id`	`<widget>.<scenario>.<short-name>`	Unique across the catalogue
`widget`, `scenario`, `decision`	Surface; the scenario in `KARAKTER.md` §4 (`S…`); the decision point in §3 (`K…` / `H…`)	Matches the file name and the documents
`when.all`	List of conditions `{var, band}`; unconditional if empty	`var`: `domain.O/C/E/A/S`, `facet.S1…C6`, `rhythm`, `interest.top`, `quality`; forbidden: `motif`, `temperament`, `archetype`, `tendency`, `value`, `values`, `code` and, with form v5, `self`, `expression`, `agreement` (§6.4.2)
`effect`	`type` (`copy`, `default`, `timing`, `frequency`, `order`, `visual`), `key`, `variant`, `value`	`key` is a single-source text/setting key; `value` may be plan-relative (`wake+30m`, `bed-3h`)
`copy`	`tr` (source text), `why` (one-sentence rationale + source key)	`tr` is required if `type = copy` (except for `no_change` records); the counterparts in the other four languages are kept in a separate file, and where there is none the Turkish remains (ADR-046 §8)
`evidence`	`grade`, `sources`, `note`	`sources` keys come from the reference list in §8 of the evidence map
`status`	`rule`, `ab_candidate`, `no_change`	For `rule` the grade must be strong or moderate

Note. Source: `Back-end/database/data/personality_rules/README.md`; validation in `Back-end/app/Services/PersonalityRules/RuleCatalog.php`.

The sleep example given in §5.5 of the evidence map shows the schema's use of plan-relative values. The record `sleepPlan.S1.rhythm-shift`, on the condition of Evening type and quality `ok`, gives the key `sleep.plan.default` the value `{"bed": "00:30", "wake": "08:15", "maxShiftMin": 90, "overrideIf": "sleep_nights>=14"}`; the `overrideIf` field writes the second item of the conflict order into the record itself.

6.7.2 Catalogue statistics

We recounted the catalogue files on 22 November 2026. Across seventeen files there are 341 records in total: 137 `rule` (88 unconditional, 49 conditional), 70 `ab_candidate` and 134 `no_change`. Each of the 229 scenarios is represented by at least one record. When the catalogue was first written (23 October 2026) it carried 361 records in fifteen files (134 `rule`, 99 of them unconditional; 75 `ab_candidate`; 152 `no_change`). The difference comes from the change in the goal surfaces: the 53-record `quarterFocus.json` was deleted on 24 October and `lifeTree.json` (18 records) and `sprint.json` (9 records) were written in its place (ADR-035); on 25 October `calendar.json` (6 records) was added (ADR-038). $361 - 53 + 33 = 341$. The other fourteen files stand as they were on the first day. Table 6.12 gives today's distribution per surface.

Table 6.12. Distribution of the rule catalogue per surface.

Surface	Records	`rule`	unconditional	conditional	`ab_candidate`	`no_change`	Scenarios
lifeTree	18	12	1	11	5	1	9
sprint	9	4	3	1	2	3	8

Surface	Records	rule	unconditional	conditional	ab_candidate	no_change	Scenarios
calendar	6	4	0	4	1	1	5
bodyScan	20	8	5	3	4	8	15
weightTrend	27	14	9	5	6	7	14
energyNeeds	19	6	4	2	3	10	14
nutritionPlan	23	5	4	1	8	10	18
waterTracker	24	12	9	3	5	7	16
stepCounter	22	9	7	2	3	10	15
workoutPlan	24	9	6	3	5	10	18
sleepPlan	21	9	5	4	5	7	12
pillBox	16	4	4	0	3	9	14
sync	10	4	3	1	2	4	8
budget	34	12	9	3	8	14	20
vault	18	5	4	1	1	12	15
market	21	7	6	1	3	11	16
celebrationNotification	29	13	9	4	6	10	12
Total	**341**	**137**	**88**	**49**	**70**	**134**	**229**

Note. Counted on 22 November 2026 from the files `Back-end/database/data/personality_rules/*.json`; the files have not changed since 25 October 2026. "Unconditional" denotes `rule` records whose `when.all` is empty.

The distribution bears the expected trace of the policy. 39% of the records (134/341) are "no change" records; 128 of these have the evidence grade "none". Of the `rule` records, 108 are graded moderate and 29 strong; of the A/B candidates, 49 are graded weak, 11 moderate and 10 "none". By effect type, 83 of the `rule` records are of type `copy`, 23 `visual` and 22 `timing`; each of the types `default`, `frequency` and `order` has only 3 rules. The rules predominantly change language and appearance; most of the records that change a default or a frequency (42 `default` and 15 `frequency` among `no_change`; 16 `default` and 8 `frequency` among the A/B candidates) are no change records or A/B candidates. This is the counterpart in the catalogue of the verdict that suggesting different values by personality is without evidence.

The variables on which the conditional records rest also reflect the verdict. Of the 49 conditional rules, 36 are the attenuation and time rules permitted by the evidence map: `domain.S = low` (15), `facet.S3 = low` (8), `facet.S1 = low` (5) and `rhythm` (8; 4 Evening type, 2 Morning type, 2 Intermediate type). The remaining 13 are the suggestion copy defined by the ADR-033 addendum (§6.6.1) and are in the Life Tree and Calendar: `interest.top` (6; one record for each letter), `rhythm` (5; 2 Evening type, 2 Morning type, 1 Intermediate type) and `domain.C = low` (2). No conditional rule depends on the `A`, `O` or `E` domain; these variables appear only in A/B candidates. The most frequent A/B condition is `facet.C5 = low` (11 records); no record that depends on C5 has `rule` status. That reminder frequency × C5 is the most critical open product question (§6.9) is consistent with this number. Status and condition do not always coincide: 5 A/B candidates are unconditional (suggestions that are independent of personality but await measurement, e.g. `sprint.S1.fresh-start`) and 2 `no_change` records carry a condition (e.g. `sprint.S3.highC-kept`).

6.7.3 Engine

The engine consists of three classes (`Back-end/app/Services/PersonalityRules/`; the fourth class in the same folder, `DomainMatrix`, belongs to the five-branch matrix of Chapter 7 and takes no part in rule selection). `PersonalitySignals` produces only the permitted variables from the user's stored profile: domain bands from the percentile where there is one (40–60 `mid`) and otherwise from Pusula's band; the first letter of the interest code; the quality level; and, only in a stage 2 profile, facet bands and Rhythm. Motif, temperament, archetype, tendency and work value are deliberately absent from this class. `RuleCatalog` loads the JSON files and checks the contract (`errors()`): that each file belongs to a configured surface, the required fields, the format and uniqueness of identifiers, the validity of condition variables and bands, the forbidden variables, the requirement of text for the `copy` type and the requirement of a strong or moderate grade for `rule` status. `PersonalityRuleEngine` makes the selection: `no_change` records are never returned; `ab_candidate` records are closed unless they are in the `config('personality_rules.enabled_ab')` list (today the list is empty); an unconditional rule is returned to everyone, including the user who has not taken the test; a conditional rule is selected only while quality is `ok` and when all its conditions are satisfied. If more than one rule is selected for the same `effect.key`, the conditional one wins.

The engine does not apply the rule, it only selects it; showing the text or suggesting the default is the job of the surface. The selection result is returned from the `GET /api/v1/personality/rules?widget=` endpoint (`PersonalityRuleController`); the endpoint requires a session, restricts the `widget` parameter to the seventeen surfaces in the configuration (`config('personality_rules.surfaces')`) and gives in the response, together with the rules, the signals that were used. The evidence note and internal fields do not go to the client; the source keys and the grade do. For a rule that carries text, the response gives, alongside the Turkish source text (`copy.tr`), its counterpart in the request language (`copy.text`); if there is no counterpart the Turkish text is returned, and the rationale sentence is not translated (ADR-046 §8).

6.7.4 Front-end hook

In the front end the `usePersonalityRules(widget)` hook (`Front-end/src/lib/usePersonalityRules.ts`) makes a single request per surface, shares the result for the lifetime of the tab and provides five query functions: `get`, `copy`, `value`, `variant`, `why`. A widget asks by key at a decision point; for example `rules.copy("weight.plateau.note") ?? defaultText`. If there is no rule (test not taken, quality low, request failed) every query returns `undefined` and the widget keeps its present behaviour. When stage 2 is completed or the test is retaken, `invalidatePersonalityRules()` empties the cache. This design guarantees in code the "baseline state" rule of the inventory: the personality layer only sits on top of a default and never replaces the default as a mandatory dependency. In accordance with ADR-033 §6, connecting the hook to the widgets is separate work. Today two widgets call the hook: the Life Tree reads the empty-branch invitation from the key `lifeTree.root.<branch>.invite` (`Front-end/src/widgets/goals/lifeTree/LifeTreeWidget.tsx`), and Calendar takes the hints in the habit composer from keys such as `calendar.composer.minimum.hint` (`Front-end/src/widgets/tasks/calendar/CalendarWidget.tsx`). The first fourteen surfaces of the inventory (health, finance, celebration and notification) and Sprint are not yet connected to the hook; the rules of these surfaces are returned from the endpoint, but no component reads them yet. The stage 2 flow empties the cache when the test ends (`Front-end/src/components/test/StageTwoFlow.tsx`).

6.7.5 Tests

The decisions are protected by tests. `PersonalityRuleCatalogTest` tests four things: the catalogue conforms to the contract (`errors()` is empty); every configured surface has a catalogue file; every source key is found in the reference list of the evidence map; the scenario of every record occurs in the relevant `KARAKTER.md` file and every scenario in the files is covered by at least one record. `PersonalityRuleEngineTest` tests the behaviour of the engine: that only unconditional rules are returned to a user without a test; that at low `S` the conditional rule takes the place of the unconditional one and that `no_change` records and closed A/B candidates are not returned; that stage 2 opens the facet and Rhythm rules; that a neck-and-neck percentile triggers no rule; that `caution` and `uninterpretable` quality close personality-dependent rules; that the single-flag `note` level does not close personality rules; that an A/B candidate is returned only once it has been opened; that an unknown surface is rejected and a session is required; that the signals do not contain motif, temperament or values. Finally, the ADR-033 test in `InvariantsTest` confirms that a sample record tied to temperament produces a "forbidden condition variable" error in the validator and that no record in the catalogue is tied to a forbidden variable (`Back-end/tests/Feature/InvariantsTest.php`); the ADR-044 test in the same file tests that the self-view, expression profile and agreement fields (`self`, `expression`, `agreement`) are on the forbidden list. In this way the rejection decision of §6.4 ceases to be a sentence in a document and becomes a check that runs on every change.

6.8 Sample profile case

6.8.1 Profile

To make the behaviour of the engine concrete we used the sample report profile in the front end (`Front-end/src/components/test/report/ornek-rapor.json`). The profile is at the full report tier (stage 2), its quality is `ok`, and the percentiles are computed against the published norm of the IPIP-NEO-120 (ADR-082; in the file `norms.n` = 619,150, `norms.source` = `ipip_neo120_johnson_2014`). In version 1.1 of this work the percentiles of the same profile were relative to the 683-person reference group; when the norm changed, the percentiles and the band of three facets changed (S3 and C6 from high to neck-and-neck, O5 from high to low). The profile is synthetic and extreme; under the new norm four of its domains fall to the 1st percentile (ADR-082, "Known limitations"). All five of the five domains are in the low band; Rhythm is Evening type. Table 6.13 shows the variables that the engine reads and does not read from this profile.

Table 6.13. Variables of the sample profile that reach and do not reach the rule engine.

Variable	Value	Engine signal
`O`, `C`, `E`, `A`, `S` (percentile)	1, 1, 1, 1, 12	all five `low`
S1 Managing worry	12th percentile	`low`
S3 Mood balance	60th percentile	`mid` (neck-and-neck; `high` requires above 60)
S5 Impulse control, C5 Self-discipline	8th and 1st percentile	`low` (a condition only in A/B and no change records)
C6 Deliberation, O5 Curiosity about ideas	46th and 34th percentile	`mid` and `low`

Variable	Value	Engine signal
O6 Flexibility in values	no percentile (ADR-082)	Pusula's band: `low`
Rhythm	Evening type	`evening`
Interest code	REC	`interest.top = R`
Quality	`ok`	`ok`
Mind motif, temperament	ISTP "Practical Fixer", SP "Learns by doing"	not read
Self type, work values	Quiet Observer (44%); Relationships and Achievement leading	not read
Self-view, expression profile, agreement (form v5)	`profile.self`, `profile.expression`, `profile.agreement`	not read (ADR-044 §6)

Note. The values are taken from the `norms` and `profile` fields of the version of the sample report file dated 22 November 2026.

Two features of the profile are instructive. First, within the `S` band, which is low at domain level, the facets diverge: S1 is low, whereas S3 is exactly at the threshold, in the neck-and-neck band. For this reason none of the eight conditional rules tied to S3 is triggered; S3 rules such as the plain celebration reach this user not through S3 but, where available, through S1. Second, C5 (1st percentile), one of the most extreme values of the profile, triggers no `rule`, because all the records tied to C5 are A/B candidates and are closed. The "frequent reminders at low Conscientiousness" example of ADR-025 §7 would have been written for exactly such a user; under the policy this user sees no difference in reminder frequency.

6.8.2 Rule counts

Table 6.14 compares the numbers of rules returned to a user who has not taken the test and to the sample profile. The numbers were computed on 22 November 2026 by re-implementing the selection logic of the engine (`PersonalitySignals::from`, `PersonalityRuleEngine::select`) in a separate script over today's catalogue and today's sample report file; the endpoint and the test suite were not run. In version 1.1 of this work, with the 361-record catalogue and the old percentiles, the totals of the same table were 99, 113 and 23.

Table 6.14. Number of rules selected per surface for a user without a test and for the sample profile.

Surface	No test	Sample profile	Personality-dependent rules
lifeTree	1	5	`S1.career-R`, `S2.health-evening`, `S3.lowC-small`, `S5.flat-gentle`
sprint	3	4	`S3.behind-gentle`
calendar	0	3	`S1.lowC-minimum`, `S2.lowS-miss`, `S3.evening`
bodyScan	5	7	`S6.badges-compact-low-s`, `S10.disclaimer-gentle-low-s`
weightTrend	9	11	`S1.trend-first`, `S2.toast-trend-delta`, `S5.plateau-neutral`, `S10.range-collapsed-low-s`
energyNeeds	4	4	`S4.floor-gentle-low-s1`, `S7.uncertainty-low-s1`
nutritionPlan	4	5	`S5.over-notice-gentle`
waterTracker	9	9	`S9.next-first-lowS`

Surface	No test	Sample profile	Personality-dependent rules
stepCounter	7	7	`S6.no-clock-lowS`
workoutPlan	6	7	`S7.rhythm-evening`
sleepPlan	5	6	`S1.rhythm-shift`, `S8.sjl-threshold-evening`
pillBox	4	4	—
sync	3	3	`S6.automation-evening`
budget	9	11	`S1.no-extra-nudge`, `S14.pct-second-low-s`
vault	4	5	`S3.total-first-low-s`
market	6	7	`S1.trend-first`
celebrationNotification	9	11	`S1.quiet-low-s1`, `S5.calm-low-s1`
Total	**88**	**109**	**29 rules**

Note. The surface prefix has been removed from the identifiers (e.g. `weightTrend.S1.trend-first`). It has been confirmed that all the identifiers are present in the catalogue with `status = rule`.

The 88 rules returned to the user without a test are the whole of the 88 unconditional rules in the catalogue. For the sample profile 29 personality-dependent rules are selected: 15 are conditional on `domain.S = low`, 5 on `facet.S1 = low`, 6 on `rhythm = evening`, 2 on `domain.C = low` and 1 on `interest.top = R`. That the total is 109 and not 88 + 29 = 117 shows the working of the fourth item of the conflict order: 8 of the 29 rules take the place of an unconditional rule on the same `effect.key`, and 21 add a new key. This is why the number does not change in Energy Needs, Water Tracking, Steps and sync: in Steps, for example, the unconditional text "1,800 steps left to the goal." (`stepCounter.S6.remaining-copy`) is replaced at low `S` by the text "6,200 steps today · 1,800 left to the goal." on the same key (`stepCounter.S6.no-clock-lowS`); the number remaining is visible in both versions, and only the clock context, which carries time pressure, is dropped.

6.8.3 Nature of the rules

Of the 29 rules selected, 24 fit the definition of attenuation in §6.6.1: 20 are low `S` and low S1 softenings and 4 are plan-relative Rhythm rules. The remaining 5 are the suggestion copy of the ADR-033 addendum. The rules fall into four groups.

Language softenings. In Sprint, instead of the "behind" label, the state and the time remaining are reported: "12 points below expected · 67 days left" (`sprint.S3.behind-gentle`). In the Life Tree the verdict on a goal that is running flat turns into a non-evaluative statement of state: "No change in the last two weeks" (`lifeTree.S5.flat-gentle`); in Calendar, for a day that was not checked off, the text "Today stayed empty; tomorrow is a new day." appears (`calendar.S2.lowS-miss`). In Weight Tracking the plateau text switches to a version without blame and suggests a single action: "The trend has been flat for a while. This is seen in most people losing weight around month 6 and is usually the sum of small daily deviations; that is how the body works. The next step: fix one meal time, not your goal." (`weightTrend.S5.plateau-neutral`). The entry confirmation gives the trend difference instead of the raw difference: "72.4 kg saved · Trend down 0.2 kg this week" (`weightTrend.S2.toast-trend-delta`). In the nutrition plan the "over" label turns into range language: "2,180 / 2,000 kcal · slightly above your range" (`nutritionPlan.S5.over-notice-gentle`). In Energy Needs, because S1 is low, the hedged

expression of uncertainty is turned into a number and a single step: "Your real expenditure is between 1,980 and 2,420 kcal. After three weeks you can look at your weight trend and adjust the target to it." (`energyNeeds.S7.uncertainty-low-s1`). The Body Scan disclaimer gives the same information in a calm tone: "This is general information, not a diagnosis. If there is pain or numbness, a physiotherapist can take a look." (`bodyScan.S10.disclaimer-gentle-low-s`).

Order and visual emphasis. In Weight Tracking the main number becomes the trend and the raw weigh-in moves down to the second line (`weightTrend.S1.trend-first`); the healthy range line is collapsed and opens with a single tap (`weightTrend.S10.range-collapsed-low-s`). In Market the 30-day direction comes first, the daily percentage moves to the second line and a fall is shown in an amber tone (`market.S1.trend-first`); in Vault the group card states the value first and the difference on the second line (`vault.S3.total-first-low-s`); in Budget, in the period comparison, "left to the goal" comes first and the percentage comparison moves down to the second line (`budget.S14.pct-second-low-s`). In the days after a budget overrun no additional "look at your budget" call is produced, but the number and the way out remain in the widget unchanged (`budget.S1.no-extra-nudge`). The celebration switches to the quiet variant, and the unread-notification badge from saturated red to a neutral tone (`celebrationNotification.S1.quiet-low-s1`, `celebrationNotification.S5.calm-low-s1`). None of these rules removes a number.

Plan-relative times. The Evening type Rhythm opens four time and threshold rules. The initial default of the sleep plan shifts to 00:30–08:15 (`sleepPlan.S1.rhythm-shift`); the workout suggestion is placed between wake +9 and +11 hours, at the latest three hours before bedtime, and with this plan falls in the range of approximately 17:15–19:15 (`workoutPlan.S7.rhythm-evening`); in sync automation the sleep data is taken 90 minutes after waking (`sync.S6.automation-evening`); for social jetlag the "marked" threshold is kept high (`sleepPlan.S8.sjl-threshold-evening`). The first three rules carry the field `overrideIf: sleep_nights>=14` in the record and yield to measured behaviour once at least 14 nights have been measured; the threshold rule does not have this field.

Suggestion copy. The profile receives five records of the kind defined by the ADR-033 addendum. In the Life Tree the invitation of the empty career branch is written according to the first letter of the interest code: "The realistic side of your interests leads: a first goal you make with your hands, with a concrete output — build, set up, or repair something." (`lifeTree.S1.career-R`); the invitation of the empty health branch comes according to Rhythm: "You are an evening person: instead of an early morning workout, a late-afternoon or evening goal sticks better for you." (`lifeTree.S2.health-evening`). The low `C` band suggests a small start in two composers: in the Life Tree "Start small and precise: the commitment type first, at most two key results; let one be tied to a habit." (`lifeTree.S3.lowC-small`), and in Calendar the filled-in example of the smallest-version field "The two-minute version: one page, one lap, one sentence" (`calendar.S1.lowC-minimum`). The time example of Calendar also changes according to Rhythm: "Evening 20:00, after dinner" (`calendar.S3.evening`). These records are additions, not softenings; none of them changes a value or a limit. One of them is in tension with the evidence map: the expression "sticks better for you" in the health invitation carries an adherence claim, whereas it has not been shown that a chronotype-aligned time increases adherence (§6.3.3, §6.5.3). The unconditional text of the workout suggestion deliberately avoids this claim; the invitation text needs to be reviewed by the same standard.

The case shows how the policy is to be read by looking at one user. A profile that is low in all five domains and whose self-discipline is at the 1st percentile may, in marketing language, look like a user who "needs to be treated very differently". The engine gives this user 80 of the 109 rules the same as everyone else; of the remaining 29 rules, 24 only soften the tone, change the order or shift the times according to a sleep plan, and 5 add suggestion copy to an empty area. The goal, the ratio, the step size and the number of reminders do not change.

6.9 Limitations and open gaps

The character trace was designed to personalise as much as the evidence can bear; the limits of this design are therefore the limits of the evidence. Table 6.15 summarises the open gaps and the ways of closing them; the five most important are then discussed.

Table 6.15. Open gaps and the way of closing them.

Gap	Present state	Way of closing
Language × personality interaction	Main effect strong; interaction tested nowhere	`copy_variant_shown` → action within 24 h; opening latency after a bad state, ABAB
Reminder frequency × C5	Searched for in five notes, not found; 11 A/B records are waiting	The `reminder_*` chain, frequency variants in ABAB
A default by chronotype increasing adherence	No study; the RCT excluded intermediate types	`default_changed` and plan adherence, by Rhythm class
Streak × personality × dropout	No personality moderator; the ADR-004 tension was resolved by the addendum of 25 October 2026 (retroactive check-in, "streak repair"), its effect has not been measured	7-day check-in after a streak break, days with and without repair used
Celebration intensity × `E`/E6	No adult sample; proxy mapping of findings from children	Action after `celebration_*`
Evidence from Türkiye	A single direct personality × persuasion study, at the level of perception	In-product measurement; local longitudinal data
Primary values that could not be verified	Some effect sizes only from the abstract or from secondary reporting	Confirmation from the primary source before citing
Product prerequisites	The notification channel covers medication only; events are not recorded; texts are in two places	Infrastructure work (below)

Note. Source: `docs/research/reports/Karakter izi kanıt haritası.md` §7; ADR-033 "Impact and open work".

The language × personality interaction has not been tested. This is the most fundamental limitation of the chapter. The assumptions that the "behind" or "exceeded" label produces dropout at low `S`, that loss framing is more harmful at low `S` and that the self-compassion tone works better at low `S` are all a combination of a main effect and logical inference. The policy's response to this is to keep personality-dependent rules in the direction of attenuation only: even if there is no interaction, softened language is a calmer version of the principles that the unconditional rule already applies to everyone, and no harm is expected. But this response does not show that the rules provide a *benefit*; it argues only that the harm is limited. A claim of benefit can be made only after the measurement plan has been in operation.

Reminder frequency × C5. The inventory's roughly 60 frequency and timing scenarios, the first example of ADR-025 §7 and the most populous A/B condition in the catalogue all depend on the same question: does a user with low self-discipline benefit from more frequent reminders? The only direct study available reports no difference (Dennis et al., 2015); the finding that frequent notifications do not lower engagement (Morrison et al., 2017) is, for its part, independent of personality. This question remains the product's most critical open question.

Evidence from Türkiye. The only direct Turkish study on personality × persuasion is at the level of perception (Alkış & Taşkaya Temizel, 2015). No longitudinal data in a Turkish sample could be found between the Big Five and BMI, weighing, budget or goal commitment; a Turkish reliability study of the MEQ exists (Pündük et al., 2005), but there is no national adult chronotype distribution, and the available percentages come from student samples. Local data on saving behaviour (ING Türkiye, 2023) support only cultural defaults (e.g. gram gold in the empty Vault example). For this reason all effect sizes in this chapter carry an assumption of generalisability when transferred to the Turkish user.

The ADR-004 tension. The evidence map had asked that the habit streak surface not be opened until the tension between the broken-streak evidence (Silverman & Barasch, 2023) and the "no grace day" decision had been resolved; this was the only point at which the evidence map directly conflicted with an existing decision. In accordance with the project rule the ADR was updated first: the addendum to ADR-004 dated 25 October 2026 (ADR-038) counts a retroactive check-in towards the streak without adding a tolerance day, and the Calendar widget was opened together with this addendum. What remains open is measuring in the product whether repair actually protects continuation; the addendum envisages narrowing the window if the share of retroactive check-ins exceeds 15%.

Prerequisites. Applying the rules to the surfaces depends on infrastructure work: opening the reminder channel beyond medication; recording the events `widget_view`, `reminder_*`, `default_changed`, `copy_variant_shown` and `rule_overridden`; a table for user pinning; moving celebration and notification texts to a single source; rendering the `SyncConnect` component. The order of implementation suggested by the inventory file follows the strength of the evidence: first the unconditional rules (no loss language, framing rule, no broken streak, amber fall), then the low `S` softenings, and last the times derived from Rhythm (`Front-end/src/widgets/KARAKTER-IZI.md`, "Next step").

Finally, every personality-dependent rule in this chapter depends on the outcome of the psychometric validation of the Turkish version of the test. The test is built on open-source instruments that have been published and tested in large samples; the validation of the Turkish version is continuing with the pilot (Chapter 4). If the results of Gates 1–4 show facet reliability to be lower than expected, it may be necessary to widen the neck-and-neck band or to reduce the facet-conditional rules to the domain band. The unconditional rules are not affected by this outcome; the part of the character trace that is most robust in terms of evidence is the part that is independent of personality.

Chapter references

Ajzen, I., Czasch, C., & Flood, M. G. (2009). From intentions to behavior: Implementation intention, commitment, and conscientiousness. *Journal of Applied Social Psychology, 39*(6), 1356–1372. https://doi.org/10.1111/j.1559-1816.2009.00485.x

Albertson, E. R., Neff, K. D., & Dill-Shackleford, K. E. (2015). Self-compassion and body dissatisfaction in women: A randomized controlled trial. *Mindfulness, 6*(3), 444–454. https://mijn.bsl.nl/self-compassion-and-body-dissatisfaction-in-women-a-randomized-c/653090

Alkış, N., & Taşkaya Temizel, T. (2015). The impact of individual differences on influence strategies. *Personality and Individual Differences, 87*, 147–152. https://open.metu.edu.tr/handle/11511/30500

Allen, M. S., & Robson, D. A. (2020). Personality and body dissatisfaction: An updated systematic review with meta-analysis. *Body Image, 33*, 77–89. https://doi.org/10.1016/j.bodyim.2020.02.001

Alslaity, A., Chan, G., & Orji, R. (2023). A panoramic view of personalization based on individual differences in persuasive and behavior change interventions. *Frontiers in Artificial Intelligence, 6*, 1125191. https://doi.org/10.3389/frai.2023.1125191

Bazley, W. J., Cronqvist, H., & Mormann, M. (2021). Visual finance: The pervasive effects of red on investor behavior. *Management Science, 67*(9), 5616–5641. https://papers.ssrn.com/ab_id=2992812

Benn, Y., Webb, T. L., Chang, B. P. I., & Harkin, B. (2016). What is the psychological impact of self-weighing? A meta-analysis. *Health Psychology Review, 10*(2), 187–203. https://doi.org/10.1080/17437199.2016.1138871

Bidargaddi, N., et al. (2018). To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. *JMIR mHealth and uHealth, 6*(11), e10123. https://doi.org/10.2196/10123

Billieux, J., Rochat, L., Rebetez, M. M. L., & Van der Linden, M. (2008). Are all facets of impulsivity related to self-reported compulsive buying behavior? *Personality and Individual Differences, 44*(6), 1432–1442. https://doi.org/10.1016/j.paid.2007.12.011

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin, 130*(6), 887–919. https://doi.org/10.1037/0033-2909.130.6.887

Breines, J. G., & Chen, S. (2012). Self-compassion increases self-improvement motivation. *Personality and Social Psychology Bulletin, 38*(9), 1133–1143. https://pubmed.ncbi.nlm.nih.gov/22645164/

Brown, S., & Taylor, K. (2014). Household finances and the "Big Five" personality traits. *Journal of Economic Psychology, 45*, 197–212. https://eprints.whiterose.ac.uk/82777/3/WRRO_82777.pdf

Brummelman, E., et al. (2014). "That's not just beautiful—That's incredibly beautiful!" The adverse impact of inflated praise on children with low self-esteem. *Psychological Science, 25*(3), 728–735. https://doi.org/10.1177/0956797613514251

Carlin, B., Olafsson, A., & Pagel, M. (2023). Mobile apps and financial decision making. *Review of Finance, 27*(3), 977–996. https://doi.org/10.1093/rof/rfac040

Churchill, S., & Jessop, D. C. (2011). Too impulsive for implementation intentions? Evidence that impulsivity moderates the effectiveness of an implementation intention intervention. *Psychology & Health, 26*(5), 517–530. https://doi.org/10.1080/08870441003611536

Cianci, A. M., Klein, H. J., & Seijts, G. H. (2010). The effect of negative feedback on tension and subsequent performance: The main and interactive effects of goal content and conscientiousness. *Journal of Applied Psychology, 95*(4), 618–630. https://doi.org/10.1037/a0019130

Ciocarlan, A., Masthoff, J., & Oren, N. (2019). Actual persuasiveness: Impact of personality, age and gender on message type susceptibility. *PERSUASIVE 2019, Lecture Notes in Computer Science, 11433*, 283–294. https://link.springer.com/10.1007/978-3-030-17287-9_23

Covey, J. (2014). The role of dispositional factors in moderating message framing effects. *Health Psychology, 33*(1), 52–65. https://durham-repository.worktribe.com/output/1477215

D'Acunto, F., et al. (2019). *Status app peer-spending feedback* [Chicago Booth Review summary]. https://chicagobooth.edu/review/how-crowdsourcing-app-can-stop-people-overspending (reference to be verified)

Dai, H., Milkman, K. L., & Riis, J. (2014). The fresh start effect. *Management Science, 60*(10), 2563–2582. https://faculty.wharton.upenn.edu/wp-content/uploads/2014/06/Dai_Fresh_Start_2014_Mgmt_Sci.pdf

Dalton, A. N., & Spiller, S. A. (2012). Too much of a good thing: The benefits of implementation intentions depend on the number of goals. *Journal of Consumer Research, 39*(3), 600–614. https://ideas.repec.org/a/oup/jconrs/doi10.1086-664500.html

Dennis, M., Masthoff, J., & Tintarev, N. (2015). *Adapting reminders to conscientiousness* [PATH 2015 workshop paper]. https://eprints.bournemouth.ac.uk/23268/1/PATH2015-paper6.pdf (reference to be verified)

Donnelly, G., Iyer, R., & Howell, R. T. (2012). The Big Five personality traits, material values, and financial well-being of self-described money managers. *Journal of Economic Psychology, 33*(6), 1129–1142. https://ideas.repec.org/a/eee/joepsy/v33y2012i6p1129-1142.html

Dorison, C. A., et al. (2022). In COVID-19 health messaging, loss framing increases anxiety with little-to-no concomitant benefits: Experimental evidence from 84 countries. *Affective Science, 3*(3), 577–602. https://doi.org/10.1007/s42761-022-00128-3

Druiven, S. J. M., et al. (2020). Stability of chronotype over a 7-year follow-up period and its association with severity of depressive and anxiety symptoms. *Depression and Anxiety, 37*(5), 466–474. https://doi.org/10.1002/da.22995

Elfhag, K., & Morey, L. C. (2008). Personality traits and eating behavior in the obese. *Eating Behaviors, 9*, 285–293. https://europepmc.org/search?query=%22Personality%20traits%20and%20eating%20behavior%20in%20the%20obese%22

Facer-Childs, E., & Brandstaetter, R. (2015). The impact of circadian phenotype and time since awakening on diurnal performance in athletes. *Current Biology, 25*(4), 518–522. https://research.birmingham.ac.uk/en/publications/the-impact-of-circadian-phenotype-and-time-since-awakening-on-diu/

Fischer, D., Lombardi, D. A., Marucci-Wellman, H., & Roenneberg, T. (2017). Chronotypes in the US: Influence of age and sex. *PLoS ONE, 12*(6), e0178782. https://doi.org/10.1371/journal.pone.0178782

Fishbach, A., & Dhar, R. (2005). Goals as excuses or guides. *Journal of Consumer Research, 32*(3), 370–377. https://ideas.repec.org/a/oup/jconrs/v32y2005i3p370-377.html

Furnham, A., Moutafi, J., & Crump, J. (2003). The relationship between the revised NEO-Personality Inventory and the Myers-Briggs Type Indicator. *Social Behavior and Personality, 31*(6), 577–584. https://doi.org/10.2224/sbp.2003.31.6.577

Gallagher, K. M., & Updegraff, J. A. (2012). Health message framing effects on attitudes, intentions, and behavior: A meta-analytic review. *Annals of Behavioral Medicine, 43*(1), 101–116. https://doi.org/10.1007/s12160-011-9308-7

Gherzi, S., Egan, D., Stewart, N., Haisley, E., & Ayton, P. (2014). The meerkat effect: Personality and market returns affect investors' portfolio monitoring behaviour. *Journal of Economic Behavior & Organization, 107*, 512–526. https://openaccess.city.ac.uk/id/eprint/16280/1/Meerkat_July2014_Final.pdf

Ghosh, A. P., & Huang, L. (2020). *Dynamic budget monitoring: When access to budget feedback leads to increase in spending* [CEPR technical report]. https://cepr.org/system/files/2022-08/Dynamic%20Budget%20Monitoring%20-%20When%20Access%20to%20Budget%20Feedback%20Leads%20to%20Increase%20in%20Spending%20-%20Anastasiya%20Pocheptsova%20Ghosh%20&%20Liang%20Huang.pdf

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics, 112*(2), 631–645. https://Rady.ucsd.edu/_files/faculty-research/uri-gneezy/risk-taking.pdf

Gold, J. M., et al. (2020). Conscientiousness in weight loss maintainers and regainers. *Health Psychology, 39*(5), 421–429. https://doi.org/10.1037/hea0000846

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology, 38*, 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1

Grüner, H. P., & Rohde, K. (2022). *The value of choice* (CEPR Discussion Paper No. 17425). https://ideas.repec.org/p/cpr/ceprdp/17425.html

Halko, S., & Kientz, J. A. (2010). Personality and persuasive technology: An exploratory study on health-promoting mobile applications. *Persuasive 2010, Lecture Notes in Computer Science*, 150–161. https://faculty.washington.edu/jkientz/papers/Halko-Personality-Persuasive2010.pdf

Hansen, J.-I. C., Dik, B. J., & Zhou, S. (2008). An examination of the structure of leisure interests of college students, working-age adults, and retirees. *Journal of Counseling Psychology, 55*(2), 133–145. https://doi.org/10.1037/0022-0167.55.2.133

Harkin, B., et al. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin, 142*(2), 198–229. https://doi.org/10.1037/bul0000025

Hirsh, J. B., & Inzlicht, M. (2008). The devil you know: Neuroticism predicts neural response to uncertainty. *Psychological Science, 19*(10), 962–967. https://www.psychologicalscience.org/news/releases/uncertainty-can-be-more-stressful-than-clear-negative-feedback.html

Hirsh, J. B., Kang, S. K., & Bodenhausen, G. V. (2012). Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits. *Psychological Science, 23*(6), 578–581. https://doi.org/10.1177/0956797611436349

ING Türkiye. (2023). *Tasarruf Eğilimleri Araştırması 2023 4. çeyrek* (IPSOS) [Press report]. https://www.sozcu.com.tr/tasarruf-sahiplerinin-tercihi-altin-ve-nakit-oldu-p24545 (reference to be verified)

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy, 3*(2), 159–186. https://www.cambridge.org/core/journals/behavioural-public-policy/article/when-and-why-defaults-influence-decisions-a-metaanalysis-of-default-effects/67AF6972CFB52698A60B6BD94B70C2C0

Jia, Y., Xu, B., Karanam, Y., & Voida, S. (2016). Personality-targeted gamification: A survey study on personality traits and motivational affordances. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2001–2013. https://doi.org/10.1145/2858036.2858515

Jokela, M., et al. (2013a). Association of personality with the development and persistence of obesity: A meta-analysis based on individual-participant data. *Obesity Reviews, 14*(4), 315–323. https://doi.org/10.1111/obr.12007

Joshanloo, M. (2022). Personality trait level and change predict future financial well-being. *Personality and Individual Differences, 191*, 111575. https://flosse.dss.gov.au/flossejspui/handle/10620/18954

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87*(4), 797–807. https://doi.org/10.1037/0021-9010.87.4.797

Kappes, A., Oettingen, G., & Pak, H. (2012). Mental contrasting and the self-regulation of responding to negative feedback. *Personality and Social Psychology Bulletin, 38*(7), 845–857. https://doi.org/10.1177/0146167212446833

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science, 62*(12), 3393–3411. https://doi.org/10.1287/mnsc.2015.2296

Keller, C., & Siegrist, M. (2015). Does personality influence eating styles and food choices? Direct and indirect effects. *Appetite, 84*, 128–138. https://europepmc.org/article/MED/25308432

Keller, J., et al. (2021). Habit formation following routine-based versus time-based cue planning: A randomized controlled trial. *British Journal of Health Psychology, 26*(3), 807–824. https://doi.org/10.1111/bjhp.12504

Kelly, K. R., & Jugovic, H. (2001). Concurrent validity of the online version of the Keirsey Temperament Sorter II. *Journal of Career Assessment, 9*(1), 49–59. https://doi.org/10.1177/106907270100900104

Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The goal-gradient hypothesis resurrected. *Journal of Marketing Research, 43*(1), 39–58. https://home.uchicago.edu/ourminsky/Goal-Gradient_Illusionary_Goal_Progress.pdf

Kliger, D., & Gilad, D. (2012). Red light, green light: Color priming in financial decisions. *Journal of Socio-Economics, 41*(5), 738–745. https://ideas.repec.org/a/eee/soceco/v41y2012i5p738-745.html

Koo, M., & Fishbach, A. (2012). The small-area hypothesis: Effects of progress monitoring on goal adherence. *Journal of Consumer Research, 39*(3), 493–509. https://academic.oup.com/jcr/article/39/3/493/1795380

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit. *Personnel Psychology, 58*(2), 281–342. https://doi.org/10.1111/j.1744-6570.2005.00672.x

Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology, 40*(6), 998–1009. https://doi.org/10.1002/ejsp.674

Lipnevich, A. A., Credé, M., Hahn, E., Spinath, F. M., Roberts, R. D., & Preckel, F. (2017). How distinctive are morningness and eveningness from the Big Five factors of personality? A meta-analytic investigation. *Journal of Personality and Social Psychology, 112*(3), 491–509. https://www.ncbi.nlm.nih.gov/pubmed/27977220

Lukas, M., & Howard, R. C. (2023). The influence of budgets on consumer spending. *Journal of Consumer Research, 49*(5), 697–720. https://academic.oup.com/jcr/article/49/5/697/6603733

Martínez, G. J., et al. (2021). Predicting participant compliance with fitness tracker wearing and ecological momentary assessment protocols in information workers: Observational study. *JMIR mHealth and uHealth, 9*(11), e22218. https://doi.org/10.2196/22218

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences, 114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

Mazri, F. H., et al. (2020). The association between chronotype and dietary pattern among adults: A scoping review. *International Journal of Environmental Research and Public Health, 17*(1), 68. https://doi.org/10.3390/ijerph17010068

McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality, 57*(1), 17–40. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x

Molloy, G. J., O'Carroll, R. E., & Ferguson, E. (2014). Conscientiousness and medication adherence: A meta-analysis. *Annals of Behavioral Medicine, 47*(1), 92–101. https://doi.org/10.1007/s12160-013-9524-4

Morrison, L. G., et al. (2017). The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: An exploratory trial. *PLoS ONE, 12*(1), e0169162. https://doi.org/10.1371/journal.pone.0169162

Munro, I. A., Bore, M. R., Munro, D., & Garg, M. L. (2011). Using personality as a predictor of diet induced weight loss and weight management. *International Journal of Behavioral Nutrition and Physical Activity, 8*, 129. https://europepmc.org/article/MED/22112231

Ndulue, C., Oyebode, O., Iyer, R. S., Ganesh, A., Ahmed, S. I., & Orji, R. (2022). Personality-targeted persuasive gamified systems: Exploring the impact of application domain on the effectiveness of behaviour change strategies. *User Modeling and User-Adapted Interaction, 32*(1–2), 165–214. https://doi.or

g/10.1007/s11257-022-09319-w

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science, 7*(4), 384–403. https://europepmc.org/abstract/MED/26168474

Olafsson, A., & Pagel, M. (2017). *The ostrich in us: Selective attention to financial accounts, income, spending, and liquidity* (NBER Working Paper No. 23945). https://www.nber.org/papers/w23945

Open Heart. (2026). *Chronotype-aligned exercise timing* [Randomised controlled trial; ISRCTN14276441]. https://bmjgroup.com/timing-exercise-to-match-body-clock-chronotype-may-lower-cardiovascular-disease-risk/ (reference to be verified)

Oyibo, K., Orji, R., & Vassileva, J. (2017). Investigation of the influence of personality traits on Cialdini's persuasive strategies. *Personalization in Persuasive Technology Workshop, CEUR Workshop Proceedings, 1833*. http://ceur-ws.org/Vol-1833/4_Oyibo.pdf

Pacanowski, C. R., Linde, J. A., & Neumark-Sztainer, D. (2015). Self-weighing: Helpful or harmful for psychological well-being? *Current Obesity Reports, 4*(1), 65–72. https://europepmc.org/article/MED/26627092

Phillips, L. A., Burns, E., & Leventhal, H. (2021). Time-of-day differences in treatment-related habit strength and adherence. *Annals of Behavioral Medicine, 55*(3), 280–285. https://doi.org/10.1093/abm/kaaa042

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research, 57*(3), 210–221. https://doi.org/10.1037/1065-9293.57.3.210

Pündük, Z., Gür, H., & Ercan, İ. (2005). Sabahçıl-akşamcıl anketi Türkçe uyarlamasında güvenilirlik çalışması. *Türk Psikiyatri Dergisi, 16*(1), 40–45. https://avesis.uludag.edu.tr/yayin/99c6a668-bc64-4417-bd78-6a3a55b1d31c/a-reliability-study-of-the-turkish-version-of-the-mornings-evenings-questionnaire-sabahcil-aksamcil-anketi-turkce-uyarlamasinda-guvenilirlik-calismasi

Purcell, K., et al. (2014). The effect of rate of weight loss on long-term weight management: A randomised controlled trial. *The Lancet Diabetes & Endocrinology, 2*(12), 954–962. https://doi.org/10.1016/S2213-8587(14)70200-1

Rounds, J., & Su, R. (2014). The nature and power of interests. *Current Directions in Psychological Science, 23*(2), 98–103. https://doi.org/10.1177/0963721414522812

Shapiro, G. K., & Burchell, B. J. (2012). Measuring financial anxiety. *Journal of Neuroscience, Psychology, and Economics, 5*(2), 92–103. https://crr.bc.edu/avoidance-comes-with-financial-anxiety/

Sharp, B., Danenberg, N., & Bellman, S. (2018). Psychological targeting [Letter to the editor]. *Proceedings of the National Academy of Sciences, 115*(34), E7890. https://doi.org/10.1073/pnas.1810436115

Sicherman, N., Loewenstein, G., Seppi, D. J., & Utkus, S. P. (2016). Financial attention. *Review of Financial Studies, 29*(4), 863–897. https://doi.org/10.1093/rfs/hhv073

Silverman, J., & Barasch, A. (2023). On or off track: How (broken) streaks affect consumer decisions. *Journal of Consumer Research, 49*(6), 1095–1117. https://doi.org/10.1093/jcr/ucac029

Singh, B., Murphy, A., Maher, C., & Smith, A. E. (2024). Time to form a habit: A systematic review and meta-analysis of health behaviour habit formation and its determinants. *Healthcare, 12*(23), 2488. https://doi.org/10.3390/healthcare12232488

Stach, M., Mulansky, L., Reichert, M., Pryss, R., & Beierle, F. (2024). Call to action: Investigating interaction delay in smartphone notifications. *Sensors, 24*(8), 2612. https://europepmc.org/article/PMC/PMC11053777

Stein, R., & Swan, A. B. (2019). Evaluating the validity of Myers-Briggs Type Indicator theory: A teaching tool and window into intuitive psychology. *Social and Personality Psychology Compass, 13*(2), e12434. https://doi.org/10.1111/spc3.12434

Stoll, G., Rieger, S., Lüdtke, O., Nagengast, B., Trautwein, U., & Roberts, B. W. (2017). Vocational interests assessed at the end of high school predict life outcomes assessed 10 years later over and above IQ and Big Five personality traits. *Journal of Personality and Social Psychology, 113*(1), 167–184. https://europepmc.org/search?query=TITLE%3A%22Vocational%20interests%20assessed%20at%20the%20end%20of%20high%20school%22

Sutin, A. R., et al. (2016). The five-factor model of personality and physical inactivity: A meta-analysis of 16 samples. *Journal of Research in Personality, 63*, 22–28. https://doi.org/10.1016/j.jrp.2016.05.001

Teixeira, G. P., Mota, M. C., & Crispim, C. A. (2018). Eveningness is associated with skipping breakfast and poor nutritional intake in Brazilian undergraduate students. *Chronobiology International, 35*(3), 358–367. https://europepmc.org/search?query=%22Eveningness%20is%20associated%20with%20skipping%20breakfast%22

Thomas, D. M., et al. (2014). Effect of dietary adherence on the body weight plateau: A mathematical model incorporating intermittent compliance with energy intake prescription. *American Journal of Clinical Nutrition, 100*(3), 787–795. https://doi.org/10.3945/ajcn.113.079822

Webb, T. L., Christian, J., & Armitage, C. J. (2007). Helping students turn up for class: Implementation intentions and conscientiousness. *Learning and Individual Differences, 17*(4), 316–327. https://eprints.whiterose.ac.uk/id/eprint/10755

Westenhoefer, J., Stunkard, A. J., & Pudel, V. (1999). Validation of the flexible and rigid control dimensions of dietary restraint. *International Journal of Eating Disorders, 26*, 53–64. https://europepmc.org/article/MED/10349584

White, B. X., Jiang, D., & Albarracín, D. (2021). The limits of defaults: The influence of decision time on default effects. *Social Cognition, 39*(5), 543–569. https://doi.org/10.1521/soco.2021.39.5.543

Wilson, K. E., & Dishman, R. K. (2015). Personality and physical activity: A systematic review and meta-analysis. *Personality and Individual Differences, 72*, 230–242. https://doi.org/10.1016/j.paid.2014.08.023

Xu, Y., Beller, A. H., Roberts, B. W., & Brown, J. R. (2015). Personality and young adult financial distress. *Journal of Economic Psychology, 51*, 90–100. https://experts.illinois.edu/en/publications/personality-and-young-adult-financial-distress/

Zheng, Y., et al. (2015). Self-weighing in weight management: A systematic literature review. *Obesity, 23*(2), 256–265. https://doi.org/10.1002/oby.20946

In-repository documents. ADR-004, ADR-010, ADR-013, ADR-019, ADR-024, ADR-025, ADR-029, ADR-031, ADR-032, ADR-033, ADR-035, ADR-038, ADR-044, ADR-046, ADR-073, ADR-082 (`docs/decisions/`); `docs/research/reports/Karakter izi kanıt haritası.md`; `docs/research/research_notes/Karakter izi kanıt haritası/` (six notes); `Front-end/src/widgets/KARAKTER-IZI.md` and the seventeen `KARAKTER.md` files that have a rule catalogue; `Back-end/database/data/personality_rules/`; `Back-end/app/Services/PersonalityRules/`; `Front-end/src/lib/usePersonalityRules.ts`; `Front-end/src/components/test/report/ornek-rapor.json`.

7. The Five-Branch Character Matrix: Carrying Personality into Behaviour across Life Domains

Chapter 6 treated the transfer of the personality profile to the dashboard at the level of behaviour rules and argued that the personality-dependent layer should remain confined to rules that only soften, that can be undone and that can be measured. This chapter examines the second application of the same evidence policy: *telling* users about their own profile in the five branches of their life (career, finance, health, relationships, spiritual life). Whereas the rules of Chapter 6 change what a widget says, the object of this chapter, the **five-branch character matrix**, is a text corpus that addresses the user directly: "you in this branch" (ADR-037). The two tasks share one question: what can a Big Five band say about behaviour in a life domain, and on what evidence? The chapter covers the two-layer principle of the matrix (§7.2), the method of the five-domain literature review (§7.3), the findings branch by branch (§7.4), the pattern across branches (§7.5), the data contract of the corpus (§7.6), its transfer to the dashboard and the report (§7.7), its operation on sample profiles (§7.8) and its limits (§7.9). The first version of the corpus consists of 76 cells and 209 sources (`Backend/database/data/domain_matrix/`); the tables in this chapter are derived from that corpus and from its synthesis report (`docs/research/reports/Beş dal karakter matrisi.md`).

7.1 Problem and approach

7.1.1 From the report to the branch

The character analysis report (Chapter 5) tells users about their domain scores, facets, mind motif and temperament; this narrative, however, is independent of life domains. When a user sets a goal on the finance branch of the Life Tree (ADR-034), the report has not a single sentence to say about finance; the only links between profile and branch are the rules of Chapter 6 and the small invitations on the tree (ADR-035). The user vision asked for more: to give, systematically, the behaviours, strengths, points of caution and advice of the personality profile in the five branches; to build, for life domains, the counterpart of the "egonomi"-type narratives produced for workplace behaviour (ADR-037, "Question").

This request is exposed to all four personalisation traps of Chapter 6. When products that generate type narratives tell the user "your type relates to finance in such and such a way", they make two claims at once: that the type is a stable measurement and that this type predicts financial behaviour. Chapter 6 had shown that the first collapses in the Gate 0 simulation (Chapter 4) and in the MBTI literature (Pittenger, 2005), and that there is no published evidence for the second. ADR-037 therefore defines the difference not as "type" but as "evidence": cells are written only at the level for which effect sizes have been published, the Big Five domains and facets; motif and temperament supply the heading (ADR-037, "Rationale").

7.1.2 The limit of the "Rehnüma standard" claim

The phrase "Rehnüma standard" in the title of ADR-037 should be read not as a marketing claim but as a methodological commitment: every cell sourced, every source graded, every grade checked by contract. This commitment does not guarantee the correctness of the content. Most of the reviews rest on meta-analysis abstracts and secondary citations (§7.3.2); the percentiles on which the bands rest are relative not to a Turkish-speaking sample but to the published international sample of the IPIP-NEO-120 (619,150 people; ADR-082); the psychometric validation of the Turkish version is continuing with the pilot (ADR-029). The present content of the standard is therefore "the traceable linking of published findings tested in large samples to the user's measured band"; it has not yet been shown that the findings are reproduced in Rehnüma users (§7.9.5).

7.1.3 Three stages

The order of Chapter 6, "first the inventory, then the evidence, then the rule", was applied here as well, but the unit changed. The inventory stage fixed the five branches of the Life Tree and the behavioural questions that can be asked in each branch; the evidence stage produced five separate literature notes, which were combined in a binding synthesis report; only after this did the corpus stage write the cells and protect them with contract tests (ADR-037, "Impact and order"). The single difference is that the corpus is text, not rule: a cell does not change a widget value, it offers the user a sentence to read. For this reason it is not the "softening only" constraint of ADR-033 that applies but the "suggestion-text rule" in the addendum to the same ADR (§7.2.3).

7.2 The two-layer principle

7.2.1 The evidence layer

ADR-037 §1 divides the matrix into two layers and makes it binding that the two are not mixed. The **evidence layer** is the five domains and 30 facets of the Big Five; this is the only level for which effect sizes have been published in the literature, and the whole corpus is written in this layer. Each cell is tied to the combination of one domain (or facet) × one band (`low` / `high`); each of the behaviour, strength, caution and advice fields is sourced and carries an evidence grade. The language is not deterministic: cells state a "tendency", not an "outcome" or a "destiny" (ADR-028; report §1.5).

There is no cell for the middle band. The "neck-and-neck" rule of ADR-033 §3 holds here too: for a domain whose percentile lies in the 40–60 range the literature does not indicate a direction, and neither does the corpus. The "Five branches" section of the full report states this to the user explicitly: "In this branch your measured bands do not give a distinguishing finding; for areas where you sit in the middle, the literature does not point a direction" (`Front-end/src/components/test/report/pages/FiveDomains.tsx`).

7.2.2 The narrative layer

The **narrative layer** is the mind motif (16 codes) and the temperament (SP/SJ/NF/NT). This layer supplies the heading and the voice: at the top of the branch card stands a line such as "Systems Architect · INTJ · Seeks solutions", and below it the strengths and cautions of the motif can be read in

an expandable section; no cell, however, is selected by motif or temperament (`Back-end/app/Services/PersonalityRules/DomainMatrix.php`, `persona()`). The card says so to the user as well: "The motif is a narrative; the sentences in the tabs rest on your measured bands and sourced evidence" (`Front-end/src/widgets/goals/lifeTree/DomainInsightCard.tsx`). Where a translation from motif to domain is needed, the mapping of McCrae and Costa (1989) is used, and it is described in this way in the report too (ADR-025 "narrative only"; ADR-028).

7.2.3 The suggestion-text rule

ADR-033 had restricted the conditions under which a personality-dependent record may be a `rule` to "reducing/softening". The addendum added for the empty-branch invitation of the Life Tree (ADR-033, addendum; ADR-035) extended this definition to text records that satisfy all of five conditions: (1) the default is empty, the rule only adds text; (2) it does not change a value, remove an option or hide information; (3) the text speaks of the user's own measured trait and contains no persuasion technique (scarcity, social proof); (4) the evidence grade is at least `orta` (moderate) and the source is in the evidence map; (5) it arrives in the interface under the "Character trace" label and the "why" sentence is in the record. The rationale is that these records feed not behaviour but the user's view of themselves and, since they carry no "effect claim", do not require A/B measurement. The five-branch corpus is the design target of these five conditions: the card starts from an empty default, adds only text, speaks of the user's band, shows cells graded `orta` and above, and arrives under the label "Character trace · you in this branch" (§7.6.4, §7.7.1).

7.2.4 Why cells are not derived from the motif

The five notes reach the same verdict: for the MBTI and Keirsey there is no peer-reviewed, replicated predictive evidence beyond the Big Five (report §4). The finance note contains only two single-country self-report studies (n = 400 and n = 182) and one publisher's technical report; for Keirsey there is no financial study at all. The health note has a single doctoral dissertation and the absence of any relationship between the 16 types and exercise motivation in 157 students; there is no study for medication adherence, sleep, substance use or mortality. In the relationships note, type similarity is unrelated to satisfaction in a sample of 62 couples (Kong, 2010), and on the Big Five side too the contribution of partner similarity is below 0.5% (Dyrenforth et al., 2010; Montoya et al., 2008; Humbad et al., 2010). The career note records that the dichotomous types do not show a bimodal distribution (Bess & Harvey, 2002), that people close to the cut-off point change type on retest (Pittenger, 2005), that the theory falls short on criteria of testability (Stein & Swan, 2019) and that the prediction of effectiveness in managers is inconsistent (Gardner & Martinko, 1996). The only peer-reviewed strand in the spiritual life note is the Francis school: quest orientation is higher in intuitive (N) types (Ross & Francis, 2010), but this finding is a restatement, in the language of types, of the relationship between O5 Curiosity about ideas and the search for meaning reported by Steger et al. (2008); the samples are English and Christian, and there is no Muslim or Turkish-speaking sample.

The translation itself is also limited. In the mappings of McCrae and Costa (1989; n = 468) and Furnham (1996), the E–I axis correlates with E and the S–N axis with O at approximately .7, the J–P axis with C at approximately .5, and the T–F axis with A at only approximately .4 (report §1.4). Four consequences follow. First, **Neuroticism has no counterpart in the MBTI**; yet it is the strongest or second strongest predictor for financial distress, sleep, mental health, relationship satisfaction, job

satisfaction, burnout and meaning (§7.5). Second, the MBTI is dichotomous; the correlations were computed from continuous scores, and estimating a band from a type letter multiplies measurement error. Third, the translation covers four domains, and facet estimation is not possible. Fourth, the Keirsey temperaments (SJ ≈ low O + high C; SP ≈ low O + low C; NF ≈ high O + high A; NT ≈ high O + low A) reduce two continuous dimensions to four boxes and leave N and E out. In Rehnüma the Big Five score is always available; deriving cells from the motif would therefore mean discarding the measurement in hand and using a noisier re-measurement of it. ADR-037 records this as a rejected alternative: "writing 16×5 or 4×5 texts per motif/temperament (no evidence)".

7.3 Method of the literature review

7.3.1 The five branches and scope

The branches are the five pillars of the Life Tree (ADR-034): `career`, `finance`, `health`, `relationships`, `spiritual`. A separate note was written for each branch (`docs/research/research_notes/Beş dal karakter matrisi/{kariyer,finans,saglik,iliskiler,ruhsal-yasam}.md`; approximately 18,000 words in total). The limits of scope were drawn at the outset: the career note leaves out occupational *choice*, because Pusula-API already provides this through RIASEC and O*NET matching (Chapter 3); career behaviour (performance, procrastination, satisfaction, leadership, burnout, income trajectory) is within scope. The spiritual life note leaves anxiety and depression to the Health branch and is limited to meaning, spirituality, religiosity, practices and personal growth (ADR-034). The finance note covers saving, debt, budget tracking, impulsive spending, investment risk, financial anxiety, frequency of account checking, financial well-being and income/wealth.

7.3.2 Source selection

The five notes are separate searches (Google Scholar, PubMed/PMC, Europe PMC, institutional repositories, TR Dizin/DOAJ). Meta-analyses and multi-cohort individual-participant analyses take priority. Coefficients whose full text could not be accessed were marked in the notes as "(secondary)" or "(not verified)"; the synthesis report and the corpus preserve these labels (report §1.1, §6). The reading rule of the finance note summarises the common stance of all the notes: the findings are correlational, the effects are in the small-to-moderate band, personality does not determine any outcome on its own, and an effect size that could not be verified is left as "not reported" and no estimate is written. The corpus carries 209 unique sources (`sources.json`); the keys are the deduplicated union of the reference lists of the five notes (report §8).

7.3.3 Evidence grades

The five notes use the same four-step scale; the corpus carries three of the steps. Table 7.1 gives the definitions.

Table 7.1. Evidence grades and their reflection in the corpus.

Grade	Definition (common denominator of the five notes)	In the corpus
strong	Meta-analysis, multi-cohort individual-participant analysis or several large/longitudinal samples; consistent direction	25 cells

Grade	Definition (common denominator of the five notes)	In the corpus
moderate	A single meta-analysis, a single large panel/longitudinal study (N > ~300–500) or a few consistent primary studies	37 cells
weak	A single/small cross-sectional study, dissertation, conference abstract, secondary citation, conflicting findings or the reverse reading of a null finding	14 cells
none	No peer-reviewed evidence found	The cell is not left empty; `zayıf` (weak) + evidence note

Note. Source: report §1.2; the cell counts were counted from the `domain_matrix/*.json` files (76 cells). Under the "no empty band" rule (ADR-037 §4), a band for which no evidence exists is marked `zayıf` and explained in the `evidence.note` field; these cells are not shown on the card (§7.6.4).

7.3.4 Mapping to Pusula variables

The domains are the same as in the Pusula vocabulary (`O C E A S`); **S = Emotional stability, and high S = stable**. Findings on "high Neuroticism" in the literature are translated in the corpus as **low S**, and domain-level coefficients are read with the sign reversed (report §1.3; Chapter 6, Table 6.1). Facet codes correspond one to one with the glossary in `Front-end/src/widgets/KARAKTER-IZI.md`; the NEO facets are mapped by the following rule: S1–S6 are the reverse-coded counterparts of N1 Anxiety, N2 Angry hostility, N3 Depression, N4 Self-consciousness, N5 Impulsiveness and N6 Vulnerability (Managing worry, Managing anger, Mood balance, Social ease, Impulse control, Stress resilience); the E, O, A and C facets map directly onto their NEO names (report §1.3, table). The consequence of reverse coding at facet level is concrete: "high N5 impulsiveness → compulsive buying" in the finance note becomes `finance.S5.low` in the corpus, "N2 anger" in the relationships note becomes `relationships.S2.low`, and "N1 anxiety → search" in the spiritual life note becomes `spiritual.S1.low`. The bridge in the finance note from the UPPS impulsivity scale to the NEO facets is conceptual (Whiteside & Lynam, 2001), and the corpus states this in the evidence note of every relevant cell.

7.3.5 Points that could not be verified

Section 6 of the synthesis report collects all of the "not verified / secondary / not found" labels in the notes; §7.9 discusses them as limits. Two general points should be recorded here with regard to method. First, in none of the notes is there a randomised controlled trial that tests the claim that "a message tailored to personality changes behaviour more than an untailored one"; this confirms the first trap of Chapter 6 in all five branches. Second, the percentiles rest not on a norm of the Turkish adaptation itself but on the published international sample (ADR-082; §7.9.3), and it has been shown by simulation that the derived labels can change by 43–55% on retest (ADR-029, Gate 0); for this reason the corpus operates only on a profile of quality `ok` and outside the 40–60 percentile "neck-and-neck" band.

7.4 Findings branch by branch

The five subsections below summarise, for each branch, the core cells (5 domains × 2 bands) and the facet rows. The behaviour column is an abridged form of the corpus text; the coefficients are taken from the synthesis report and are found not in the user-facing texts of the corpus but only in the `evi-`

`dence.note` field (§7.6.2). The source column gives the main sources of the cell; the full lists are in the `domain_matrix/*.json` files.

7.4.1 Career

The career branch is the one in which the literature is densest, and seven of the ten core cells are graded `güçlü` (strong). Conscientiousness is the most consistent predictor of job performance (Barrick & Mount, 1991; Wilmot & Ones, 2019); even after the correction of Sackett et al. (2022), ρ for overall performance is at approximately .19. The strongest personality correlate of procrastination is low Conscientiousness ($\rho \approx -.6$; Steel, 2007). Emotional stability gives the most consistent relationship with job satisfaction (Judge, Heller, & Mount, 2002), burnout (Alarcon et al., 2009; Swider & Zimmerman, 2010) and career satisfaction (Ng et al., 2005). Extraversion is the most consistent correlate of leadership emergence (.31; Judge, Bono, et al., 2002), but assertiveness shows an inverted-U relationship with leadership effectiveness (Ames & Flynn, 2007). Openness is positive with learning, creativity and innovative behaviour (Hammond et al., 2011) and with career adaptability (Rudolph et al., 2017); its relationship with job satisfaction is close to zero. Agreeableness is an advantage in the team context and a small disadvantage in income and promotion (Judge et al., 2012; Ng et al., 2005). Table 7.2 summarises the cells.

Table 7.2. Career branch: domain × band × behaviour summary × evidence grade × main source.

Domain / facet	Band	Behaviour summary	Grade	Main source
C Conscientiousness	high	Task performance (ρ ≈ .19–.27), goal persistence, low procrastination, job satisfaction (.26); difficulty accepting "good enough" (evidence for a cost of excessive C is weak)	strong	Wilmot & Ones, 2019; Sackett et al., 2022; Judge, Heller, & Mount, 2002
C Conscientiousness	low	Procrastination (ρ ≈ –.6); need for external structure; implementation intentions (d = .65) and breaking work into sub-steps pay off most in this band	strong	Steel, 2007; Gollwitzer & Sheeran, 2006; Sheeran et al., 2024; Locke & Latham, 2002
S Emotional stability	high	Job and career satisfaction, low burnout, steadiness under pressure; advantage in visible roles	strong	Judge, Heller, & Mount, 2002; Alarcon et al., 2009; Ng et al., 2005
S Emotional stability	low	Job satisfaction (–.29), emotional exhaustion (≈ .5), career satisfaction (–.36); generalising negative feedback to identity; feedback directed at the person does harm	strong	Judge, Heller, & Mount, 2002; Swider & Zimmerman, 2010; Kluger & DeNisi, 1996
E Extraversion	high	Leadership emergence (.31), proactivity, promotion (.18), career satisfaction (.27); inverted-U in assertiveness	strong	Judge, Bono, et al., 2002; Ng et al., 2005; Ames & Flynn, 2007
E Extraversion	low	Small negative with promotion and career satisfaction (reverse reading); no peer-reviewed evidence for a preference for "deep work"	weak	Judge, Bono, et al., 2002; Ng et al., 2005; Grant, 2013
O Openness	high	Learning in training (.25), innovative behaviour (.2–.3), entrepreneurial intention, leadership (.24); ≈ 0 with job satisfaction	strong	Barrick & Mount, 1991; Hammond et al., 2011; Zhao et al., 2010
O Openness	low	Deepening in an established method; career adaptability low in periods of change	moderate	Rudolph et al., 2017
A Agreeableness	high	Team performance and leadership effectiveness; small negative with income and promotion, especially in men	strong	Barrick et al., 2001; Judge et al., 2012; Judge et al., 1999

Domain / facet	Band	Behaviour summary	Grade	Main source
A Agreeableness	low	Advantage in negotiation; friction in the team context; small difference in becoming an entrepreneur	moderate	Zhao & Seibert, 2006; Judge et al., 2012
C4 Achievement striving	high	The face of Conscientiousness that predicts task performance; incremental validity	moderate	Dudley et al., 2006; Judge et al., 2013
C6 Deliberation	high	The dependability face; low counterproductive work behaviour	moderate	Dudley et al., 2006; Judge et al., 2013
C1 Self-efficacy	high	Self-efficacy → performance G(r) ≈ .38; job satisfaction ≈ .26 (core self-evaluations as proxy)	strong	Judge & Bono, 2001; Stajkovic & Luthans, 1998; Colquitt et al., 2000
E3 Assertiveness	high	Inverted-U with leadership effectiveness; the middle level is the most effective region	moderate	Ames & Flynn, 2007; Grant, 2013
O5 Curiosity about ideas	high	Stronger relationship with workplace innovation than the aesthetic face	moderate	Hammond et al., 2011; DeYoung et al., 2007

Note. Source: `career.json` (10 core + 5 facet cells); coefficients from report §2.4. The career note records that it could not access the full text of most of the primary articles and that all ρ values given with `≈` need to be verified; the correction of Sackett et al. (2022) has lowered earlier validity estimates.

7.4.2 Finance

In the finance branch only two cells are `güçlü`, and both are the high bands of Conscientiousness and Emotional stability. In young adults, Conscientiousness one standard deviation higher goes together with approximately 40% higher net worth, and the effect of financial literacy appears together with Conscientiousness (Letkiewicz & Fox, 2014); time devoted to money management goes with Conscientiousness (Donnelly et al., 2012). Emotional stability is positive with subjective financial well-being (Joshanloo, 2022) and household wealth (Mosca & McCrory, 2016) and negative with financial distress (Xu et al., 2015). Extraversion raises unsecured debt along with earnings and assets (Brown & Taylor, 2014) and is associated with paying higher prices in an experimental market (Oehler et al., 2018). Openness shows a small-to-moderate relationship with earnings but is unrelated to wealth (Alderotti et al., 2023; Mosca & McCrory, 2016). No meta-analysis was found for Big Five × saving/debt; the only meta-analysis is for earnings (Alderotti et al., 2023). At facet level the only robust signal is impulsivity: UPPS "urgency" is the only robust personality predictor of compulsive buying and operates at moments of both negative and positive emotion (Billieux et al., 2008; Rose & Segrist, 2014); excitement seeking is unrelated. Table 7.3 summarises the cells.

Table 7.3. Finance branch: domain × band × behaviour summary × evidence grade × main source.

Domain / facet	Band	Behaviour summary	Grade	Main source
C Conscientiousness	high	Time on money management, saving, low debt; 1 SD ≈ 40% net worth; low risk tolerance	strong	Donnelly et al., 2012; Letkiewicz & Fox, 2014; Joshanloo, 2022
C Conscientiousness	low	Little time on money management, financial distress; literacy alone does not turn into wealth; automatic and raise-indexed saving (3.5% → 13.6%), reminder tied to an expense	moderate	Donnelly et al., 2012; Xu et al., 2015; Thaler & Benartzi, 2004; Karlan et al., 2016
S Emotional stability	high	Positive with subjective financial well-being and wealth; attention is driven by the account balance	strong	Joshanloo, 2022; Mosca & McCrory, 2016; Olafsson & Pagel, 2017

Domain / facet	Band	Behaviour summary	Grade	Main source
S Emotional stability	low	Financial distress and anxiety; information avoidance; less portfolio risk; avoidance of the red/loss frame (8 experiments); logins fall by 9.5% after a downturn	moderate	Xu et al., 2015; Shapiro & Burchell, 2012; Bazley et al., 2021; Sicherman et al., 2016
E Extraversion	high	Earnings and assets ↑ but unsecured debt ↑; overpaying, short-term intention; spending–personality fit raises satisfaction	moderate	Brown & Taylor, 2014; Oehler et al., 2018; Matz et al., 2016
E Extraversion	low	Lower unsecured debt; small negative with earnings	moderate	Brown & Taylor, 2014; Alderotti et al., 2023
O Openness	high	Small-to-moderate positive with earnings (controlling for education and cognitive ability); long-term intention; no relationship with wealth	moderate	Alderotti et al., 2023; Mosca & McCrory, 2016
O Openness	low	No distinguishing finding; small inverse relationship with earnings	weak	Alderotti et al., 2023
A Agreeableness	high	Small negative with earnings and financial satisfaction; no rationale for intervention	moderate	Alderotti et al., 2023; Mosca & McCrory, 2016
A Agreeableness	low	No finding other than a small positive with earnings	weak	Alderotti et al., 2023
S5 Impulse control	low	UPPS "urgency" is the only robust predictor of compulsive buying (n = 150 and 514); intervention evidence is weak	moderate	Billieux et al., 2008; Rose & Segrist, 2014
C5 Self-discipline	low	Correlated with compulsive buying; non-significant in regression	weak	Billieux et al., 2008
C6 Deliberation	low	Correlated with compulsive buying; non-significant in regression	weak	Billieux et al., 2008
E5 Excitement seeking	high	Unrelated to compulsive buying (null finding; no warning is derived)	moderate	Billieux et al., 2008

Note. Source: `finance.json` (10 core + 4 facet cells); coefficients from report §2.1 and §5. The C → money management path coefficient of Donnelly et al. (2012) was not verified in the note; Rustichini et al. (2016) is a secondary citation; there is no direct financial study with the 30 NEO facets.

7.4.3 Health

The health branch has the second largest number of `güçlü` cells (six core cells). Conscientiousness is the meta-analytic predictor of health behaviours (Bogg & Roberts, 2004): medication adherence (Molloy et al., 2014), lower risk of obesity (Jokela et al., 2013a) and smoking (Hakulinen et al., 2015), and longevity (Kern & Friedman, 2008; Jokela et al., 2013b). In low Conscientiousness the problem is not intensity but continuity: r = .21 with the *frequency* of activity, .06 with its *amount* (Wilson & Dishman, 2015); the implementation intention intervention worked most at low and moderate Conscientiousness (Webb et al., 2007), although this moderation was not replicated in another study (Elliott et al., 2024). Emotional stability shows an effect on sleep quality larger than that of demographics (Stephan et al., 2018; Duggan et al., 2014) and gives the strongest relationship with mental disorders (Kotov et al., 2010); this relationship halves once baseline symptoms are controlled but does not fade over time (Jeronimus et al., 2016). At low stability, emotional eating and weight fluctuation are observed (Sutin et al., 2011), but the risk of smoking and substance use is significant not with low stability alone but with the combination of low stability and low Conscientiousness (Terracciano & Costa, 2004; Turiano et al., 2012). Extraversion gives the strongest domain-level relationship of personality with physical activity (r = .23; Rhodes & Smith, 2006) and is negative with inactivity

(Sutin et al., 2016); at the same time it also comes with smoking initiation (OR 1.22; Hakulinen et al., 2015) and substance use. Agreeableness lowers the risk of alcohol use (Malouff et al., 2007), smoking and weight gain; in general health, Agreeableness, Conscientiousness and stability are the strongest trio, and the multiple R is .35 (Strickhouser et al., 2017). Openness is unrelated to smoking, weight and sleep. Table 7.4 summarises the cells.

Table 7.4. Health branch: domain × band × behaviour summary × evidence grade × main source.

Domain / facet	Band	Behaviour summary	Grade	Main source
C Conscientiousness	high	Exercise frequency, adherence to medication and plans, low obesity and smoking, good sleep hygiene, morningness, longevity; the relationships are small (r ≈ .10–.25)	strong	Bogg & Roberts, 2004; Molloy et al., 2014; Jokela et al., 2013a; Kern & Friedman, 2008
C Conscientiousness	low	A continuity problem (frequency r = .21, amount .06); adherence, weight and smoking risk; implementation intentions most in this band	strong	Wilson & Dishman, 2015; Hakulinen et al., 2015; Webb et al., 2007; Gollwitzer & Sheeran, 2006
S Emotional stability	high	Good sleep, little emotional eating, low relapse and mental health risk; the "it won't happen to me" dismissal (inference)	strong	Stephan et al., 2018; Kotov et al., 2010; Jeronimus et al., 2016
S Emotional stability	low	Stress reactivity, emotional eating, weight fluctuation, postponed sleep, self-soothing; risk when combined with low C; self-compassion r = .25 but the scale overlaps with neuroticism	strong	Sutin et al., 2011; Lahey, 2009; Terracciano & Costa, 2004; Sirois et al., 2015; Pfattheicher et al., 2017
E Extraversion	high	Physical activity (r = .23), good sleep; smoking initiation (OR 1.22), substance use; eveningness	strong	Rhodes & Smith, 2006; Sutin et al., 2016; Hakulinen et al., 2015; Lipnevich et al., 2017
E Extraversion	low	Risk of inactivity, lower sleep quality; smoking and substance risk low	strong	Rhodes & Smith, 2006; Sutin et al., 2016; Stephan et al., 2018
O Openness	high	Slightly less inactivity (OR 0.81/SD) but unrelated in meta-analyses; trying substances slightly more likely	weak	Sutin et al., 2016; Turiano et al., 2012
O Openness	low	Unrelated to smoking, weight, sleep; small, conflicting inverse tendencies	weak	Sutin et al., 2016; Terracciano & Costa, 2004
A Agreeableness	high	Low risk of alcohol use, smoking and weight gain; morningness (r = .13); one of the strongest trio in general health (R = .35)	moderate	Malouff et al., 2007; Tsaousis, 2010; Strickhouser et al., 2017
A Agreeableness	low	Alcohol/substance use and a steeper BMI increase slightly more likely; the alcohol profile is the combination of low A + low C + low S	moderate	Malouff et al., 2007; Turiano et al., 2012; Graham et al., 2017
C4 Achievement striving	high	The strongest facet relationship with activity (r = .18), −.22 with unhealthy eating; Orderliness .08, Self-control (≈ C5) .07	strong	Bogg & Roberts, 2004
C5 Self-discipline	low	Common among smokers; r = .07 with activity (the sentence "I lack discipline, I cannot exercise" is not supported)	moderate	Terracciano & Costa, 2004; Bogg & Roberts, 2004
C6 Deliberation	low	Common among smokers; marked when combined with low S (N = 1,638, cross-sectional)	moderate	Terracciano & Costa, 2004
S5 Impulse control	low	The strongest facet line with weight: the top 10% are ≈ 11 kg heavier than the bottom 10% (BLSA, N = 1,988)	moderate	Sutin et al., 2011; Gerlach et al., 2015

Domain / facet	Band	Behaviour summary	Grade	Main source
E5 Excitement seeking	high	Common among smokers; substance risk when combined with low C	mod-erate	Terracciano & Costa, 2004; Turiano et al., 2012

Note. Source: `health.json` (10 core + 5 facet cells); coefficients from report §2.2. The full texts of Graham et al. (2017) and Malouff et al. (2007) could not be read; there is no meta-analytic coefficient for E4/E5 × activity; the age moderator of the C–activity relationship and the N–mortality relationship are conflicting (report §6).

7.4.4 Relationships

In the relationships branch personality explains approximately 6% of relationship satisfaction (Dyrenforth et al., 2010); despite this small share the direction is consistent, and there are `güçlü` cells in four domains. Emotional stability gives the most consistent relationship with both one's own and the partner's satisfaction (Malouff et al., 2010; Heller et al., 2004); low stability predicts the strongest negative Big Five relationship with relationship satisfaction ($r \approx -.22$), anxious attachment ($r \approx .40$; Noftle & Shaver, 2006) and loneliness ($r = .36$; Buecker et al., 2020); the risk of separation operates mainly through satisfaction (Solomon & Jackson, 2014). Agreeableness goes with forgiveness ($r \approx .2$–$.3$; Fehr et al., 2010), empathy and compromise (Jensen-Campbell & Graziano, 2001) and exerts one of the largest partner effects on the partner's satisfaction; at low Agreeableness conflict and power tactics increase, and the observational findings of the Gottman tradition (turning toward bids for attention, soft start-up, repair) are the source of the advice (Gottman & Levenson, 2000; Driver & Gottman, 2004). Extraversion is the most consistent predictor of network size (Asendorpf & Wilpers, 1998) and gives the lowest relationship with loneliness within the Big Five ($r = -.37$; Buecker et al., 2020); at low Extraversion, brief trials of "acting extraverted" give momentary positive affect, but a fatigue cost over one week has been reported (Zelenski et al., 2012; Jacques-Hamilton et al., 2019). Conscientiousness is small positive with one's own and the partner's satisfaction ($r \approx .15$–$.20$) and negative with divorce (Roberts et al., 2007). Openness is unrelated to relationship satisfaction ($r \approx .05$–$.10$). At facet level the only consistent signal is low anger management: it is associated with behaviours in marriage that anger or upset the partner (Buss, 1991). Table 7.5 summarises the cells.

Table 7.5. Relationships branch: domain × band × behaviour summary × evidence grade × main source.

Domain / facet	Band	Behaviour summary	Grade	Main source
S Emotional stability	high	Most consistent positive with one's own and the partner's satisfaction; little loneliness and anxious attachment; being read as "doesn't care" (inference)	strong	Malouff et al., 2010; Heller et al., 2004; Buecker et al., 2020
S Emotional stability	low	r ≈ –.22 with satisfaction; anxious attachment r ≈ .40; loneliness r = .36; less forgiveness; separation through satisfaction; emotion regulation and expressing gratitude are the most direct interventions	strong	Malouff et al., 2010; Noftle & Shaver, 2006; Solomon & Jackson, 2014; Bloch et al., 2014; Lambert et al., 2010; Algoe et al., 2010
E Extraversion	high	Wide network, quick new ties; r = –.37 with loneliness; network breadth is not depth (inference)	strong	Asendorpf & Wilpers, 1998; Buecker et al., 2020; Harris & Vazire, 2016
E Extraversion	low	Loneliness and low perceived support more common on average; small relationship with avoidant attachment; scheduled one-to-one time; fatigue cost of "acting extraverted"	strong	Buecker et al., 2020; Barańczuk, 2019; Zelenski et al., 2012; Jacques-Hamilton et al., 2019; Masi et al., 2011

Domain / facet	Band	Behaviour summary	Grade	Main source
A Agreeableness	high	Little conflict, compromise, forgiveness (r ≈ .2–.3), empathy; one of the largest partner effects; postponing one's own needs (inference)	strong	Malouff et al., 2010; Fehr et al., 2010; Jensen-Campbell & Graziano, 2001
A Agreeableness	low	Conflict frequency and power tactics; negative with the partner's satisfaction; turning toward bids for attention, soft start-up, repair	moderate	Dyrenforth et al., 2010; Jensen-Campbell & Graziano, 2001; Driver & Gottman, 2004; Gottman & Levenson, 2000
C Conscientiousness	high	One's own and the partner's satisfaction (r ≈ .15–.20), negative with divorce; warmth and structure in parenting	strong	Malouff et al., 2010; Roberts et al., 2007; Prinzie et al., 2009
C Conscientiousness	low	Small positive with divorce risk; loneliness slightly more common (r = −.20, reverse reading); large partner effect	moderate	Roberts et al., 2007; Dyrenforth et al., 2010; Buecker et al., 2020
O Openness	high	Close to zero with satisfaction (r ≈ .05–.10); shared new experiences (inference)	weak	Malouff et al., 2010; Heller et al., 2004
O Openness	low	Unrelated to satisfaction; evidence neutral	weak	Malouff et al., 2010; Heller et al., 2004
S2 Managing anger	low	Behaviours that anger/upset the partner; the most consistent signal at facet level (a single primary study)	moderate	Buss, 1991
E1 Warmth	high	More positive interaction at first encounter (a single laboratory study, female dyads)	weak	Berry & Hansen, 2000

Note. Source: `relationships.json` (10 core + 2 facet cells); coefficients from report §2.3. The contribution of partner similarity is below 0.5%; the idea of a "compatible type" finds no evidence (Dyrenforth et al., 2010). The numerical breakdown of Prinzie et al. (2009) was not verified from the primary text; Big Five × relationship satisfaction in Turkish samples was not searched systematically.

7.4.5 Spiritual life

The backbone of the spiritual life branch is a single study: the study in which Steger et al. (2008) related the NEO-PI-R to the measurement of meaning (N = 275) is the sole source of nine of the ten rows at facet level. Presence of meaning gives its strongest relationships with Emotional stability (r = .35) and Conscientiousness (r = .34); the search for meaning goes with Openness (r = .20), and the relationship between search and presence of meaning is negative at low Openness (β = −.37) and close to zero at high Openness. Religiosity shows a small positive relationship with Agreeableness and Conscientiousness (Saroglou, 2010; 71 samples, N = 21,715), and this relationship is bidirectional (Entringer et al., 2023); Openness is positive with spirituality as distinct from institutional religion (Saroglou, 2002). Openness is, in addition, the strongest personality predictor of participation in adult education and is related to connection with nature (.33; Tam, 2013). At low stability, mindfulness meditation gives an effect of 0.38 on anxiety in eight weeks (47 RCTs; Goyal et al., 2014); death anxiety is high in most studies, but there is no meta-analysis and the Turkish findings are conflicting. Resilience is positive with low neuroticism and the other four domains (Oshio et al., 2018; the coefficients could not be verified). The sources the note found for the Turkish context are limited: in a Turkish-speaking sample the religious individualisation total is negative with C and A, that is, institutional religiosity is in line with the international pattern; no Turkish article was found that gives a direct "Five-Factor × religiosity scale" correlation table. Table 7.6 summarises the cells.

Table 7.6. Spiritual life branch: domain × band × behaviour summary × evidence grade × main source.

Domain / facet	Band	Behaviour summary	Grade	Main source
C Conscientiousness	high	Presence of meaning (r = .34), persistence in worship and community practice, persistence in meditation; risk of deriving meaning from productivity alone and of religious perfectionism	moderate	Steger et al., 2008; Saroglou, 2010; Entringer et al., 2023
C Conscientiousness	low	Search is independent of Conscientiousness; institutional religious attachment slightly lower; risk of abandoning meaning practices	weak	Steger et al., 2008; Saroglou, 2010
S Emotional stability	high	Presence of meaning (r = .35), resilience, "joyous exploration" curiosity; postponing deep questions (inference)	moderate	Steger et al., 2008; Oshio et al., 2018; Kashdan et al., 2018
S Emotional stability	low	Presence of meaning low; the "deficit" part of the search; death anxiety; mindfulness meditation (anxiety 0.38 / 8 weeks); no metric implying "meaningless" is shown	moderate	Steger et al., 2008; Goyal et al., 2014; Oshio et al., 2018
E Extraversion	high	Presence of meaning (r = .25; Cheerfulness .31), volunteering, one of the two domains related to belief at the within-person level	moderate	Steger et al., 2008; Carlo et al., 2005; Lavigne et al., 2013
E Extraversion	low	Presence of meaning slightly lower (small); weak with connection with nature (.14)	weak	Steger et al., 2008; Tam, 2013
O Openness	high	Search for meaning (r = .20), spirituality apart from institutions, participation in adult education (N = 10,559), connection with nature (.33), having tried meditation; the search does not reduce presence of meaning	strong	Steger et al., 2008; Saroglou, 2002; Tam, 2013; Goyal et al., 2014; McMahan & Estes, 2015
O Openness	low	Presence of meaning independent of Openness (r = .02); firm ground within tradition; when a search arises it lowers meaning (β = −.37)	moderate	Steger et al., 2008; Saroglou, 2010
A Agreeableness	high	Small positive with religiosity (r ≈ .19, not verified), presence of meaning (.21), volunteering, connection with nature (.34), benevolence values; the threshold of 120 min a week in nature (observational)	strong	Saroglou, 2010; Steger et al., 2008; Carlo et al., 2005; Roccas et al., 2002; Parks-Leduc et al., 2015; White et al., 2019; Baranski et al., 2024
A Agreeableness	low	Institutional religious attachment slightly lower; prosociality by way of Extraversion; a gratitude journal may feel artificial (inference)	weak	Carlo et al., 2005; Saroglou, 2010
C4 Achievement striving	high	The strongest C facet with presence of meaning (r = .33)	moderate	Steger et al., 2008
C3 Dutifulness	high	Presence of meaning (r = .28)	moderate	Steger et al., 2008
C2 Orderliness	high	Presence of meaning (r = .27)	moderate	Steger et al., 2008
O5 Curiosity about ideas	high	The strongest O facet with the search for meaning (r = .22; partial .28); the counterpart of the MBTI finding "intuitive type ↔ quest"	moderate	Steger et al., 2008; Ross & Francis, 2010
O6 Flexibility in values	high	The only O facet related to presence of meaning (r = .15); no evidence for spirituality	weak	Steger et al., 2008
E6 Cheerfulness	high	The strongest E facet with presence of meaning (r = .31)	moderate	Steger et al., 2008

Domain / facet	Band	Behaviour summary	Grade	Main source
E1 Warmth	high	Presence of meaning (r = .22)	mod-erate	Steger et al., 2008
A3 Helpfulness	high	The strongest A facet with presence of meaning (r = .27)	mod-erate	Steger et al., 2008
A6 Compassion	high	Search for meaning (r = .17; partial .21)	mod-erate	Steger et al., 2008
S1 Managing worry	low	Search for meaning (r = .20), which disappears when presence of meaning is controlled (partial .05): the search stems from a lack of meaning	mod-erate	Steger et al., 2008

Note. Source: `spiritual.json` (10 core + 10 facet cells); coefficients from report §2.5. The coefficients of Saroglou (2002, 2010), Parks-Leduc et al. (2015), Roccas et al. (2002) and Oshio et al. (2018) were taken from abstracts or from a secondary source (Saucier, 2019); the religiosity–A r ≈ .19 has not been verified. That the facet rows rest on a single source (N = 275) is the reason these rows cannot rise above the `orta` grade.

7.5 The pattern across branches

Read together, the five tables reveal five patterns; the synthesis report collects them in §3.

Conscientiousness and Emotional stability in almost every branch. Conscientiousness gives at least one `orta` or `güçlü` row in all five of the five branches, and its common denominator is the *continuity* of behaviour: in finance, money management and wealth (Letkiewicz & Fox, 2014); in health, activity frequency, medication adherence, obesity, smoking and longevity (Bogg & Roberts, 2004; Molloy et al., 2014; Jokela et al., 2013a; Kern & Friedman, 2008); in relationships, satisfaction and divorce (Roberts et al., 2007); in career, performance and procrastination (Wilmot & Ones, 2019; Steel, 2007); in spiritual life, presence of meaning and persistence in worship (Steger et al., 2008; Entringer et al., 2023). Emotional stability, for its part, is the carrier of *subjective outcomes*: financial well-being (Joshanloo, 2022), sleep and mental health (Stephan et al., 2018; Kotov et al., 2010), relationship satisfaction and loneliness (Malouff et al., 2010; Buecker et al., 2020), job satisfaction and burnout (Judge, Heller, & Mount, 2002; Alarcon et al., 2009), presence of meaning and resilience (Steger et al., 2008; Oshio et al., 2018). Their conjunction matters too: high Conscientiousness buffers the smoking and substance risk of low stability (Terracciano & Costa, 2004; Turiano et al., 2012). Because ADR-033 §4 prohibits writing products of variables, this interaction remains a "caution" sentence in the corpus and does not enter cell selection.

Openness: search and learning. Openness is close to zero in financial wealth, health behaviours and relationship satisfaction (Mosca & McCrory, 2016; Terracciano & Costa, 2004; Malouff et al., 2010); by contrast it is positive with the search for meaning (Steger et al., 2008), spirituality as distinct from institutional religion (Saroglou, 2002), participation in adult education, workplace innovation (Hammond et al., 2011) and earnings (Alderotti et al., 2023). At low Openness the search lowers presence of meaning; at high Openness it does not. This suggests that Openness distinguishes not behaviour but the **source of meaning**; that the Openness cells of the corpus remain `zayıf` in both bands in health and relationships and in the low band in finance (the high band of finance is `orta`), while the high Openness cell is `güçlü` in career and spiritual life, is the direct consequence of this pattern.

Agreeableness: relationships and religiosity. Agreeableness is among the strongest domains in relationship satisfaction, forgiveness (Fehr et al., 2010), empathy and the partner effect; it is small positive with religiosity in 61 countries (Saroglou, 2010; Baranski et al., 2024) and positive with volunteering and connection with nature (Carlo et al., 2005; Tam, 2013). The same domain is small negative with earnings and promotion (Alderotti et al., 2023; Ng et al., 2005; Judge et al., 2012); the advice in the career cell, "prepare the salary conversation in a market-data frame", comes from here.

Extraversion: network and income, but debt. Extraversion is positive with network size (Asendorpf & Wilpers, 1998), the reverse of loneliness (Buecker et al., 2020), physical activity (Rhodes & Smith, 2006), leadership emergence (Judge, Bono, et al., 2002), and promotion and earnings (Ng et al., 2005; Alderotti et al., 2023). At the same time it also comes with unsecured debt (Brown & Taylor, 2014), overpaying in an experimental market (Oehler et al., 2018), smoking initiation (Hakulinen et al., 2015) and substance use (Turiano et al., 2012); there is an inverted-U in assertiveness (Ames & Flynn, 2007).

Effects are small but consistent. Individual domain–behaviour correlations are mostly in the r = .10–.35 band; approximately 6% of relationship satisfaction (Dyrenforth et al., 2010), ρ ≈ .19 of overall job performance (Sackett et al., 2022) and a multiple R = .35 of general health (Strickhouser et al., 2017) are explained by personality. The personality → outcome direction is stronger than the outcome → personality direction, but the latter exists too (Mund & Neyer, 2014; Neyer & Lehnart, 2007; Entringer et al., 2023); personality change has been measured with three months of digital microtasks (Stieger et al., 2021; Roberts et al., 2017). This is the evidence for writing every cell of the corpus in the mode of "a tendency, not a destiny"; it is also the reason why the branch card never says "you are such and such".

7.6 Corpus and contract

7.6.1 Structure

The corpus is five JSON files (`Back-end/database/data/domain_matrix/{career,finance,health,relationships,spiritual}.json`) and a source dictionary (`sources.json`, 209 keys). There are 76 cells in total: 50 core cells (5 branches × 5 domains × 2 bands) and 26 facet cells (career 5, finance 4, health 5, relationships 2, spiritual life 10). Table 7.7 gives the schema.

Table 7.7. Cell schema (`domain_matrix/*.json`).

Field	Type	Content	Visible to the user?
`id`	text	`<branch>.<domain|facet>.<band>`, e.g. `finance.S5.low`; unique	no
`domain`, `trait`, `facet`, `band`	enum	branch; `O|C|E|A|S`; `S1…C6` or `null`; `low|high`	as a label
`behavior`	text	One-sentence summary of the tendency (required)	yes (fallback)
`narrative`	text	3–4 sentences: the tendency, how it appears in daily life, the other side of the coin, the closing "a tendency, not a destiny" (required)	yes ("You" tab)
`strength`, `caution`	text / `null`	Strength and caution; `null` if there is no evidence	yes ("Strengths · Cautions")

Field	Type	Content	Visible to the user?
`advice`	text / `null`	Advice (report version)	yes (fallback)
`goal_light`	text	Goal-setting light: *how* setting a goal in this branch makes it hold better (required)	yes ("When setting goals")
`evidence.grade`	enum	`güçlü|orta|zayıf`	on the branch card (the weakest grade shown)
`evidence.sources`	array	`sources.json` keys; may not be empty	on the branch card (key)
`evidence.note`	text	Coefficients, N, study type, "(inference)" labels	no (returned in the endpoint response, not rendered in the interface)

Note. Source: `DomainMatrix::errors()` and sample cells. The first schema in ADR-037 §4 (`behavior, strength, caution, advice, evidence, tr`) was extended in the third step with `narrative` and `goal_light`, and the coefficients were moved from the user text to `evidence.note` (ADR-037, "Impact and order" 3). Since 22 November 2026 the "Five branches" section of the report no longer shows the evidence grade label, the source line or the evidence note (§7.7.3).

The source language of the corpus is Turkish; the six user fields (`narrative`, `goal_light`, `behavior`, `advice`, `caution`, `strength`) have been translated into English, French, German and Arabic (`Backend/database/data/<lang>/domain_matrix/<branch>.json`; 76 cells per language). Cell selection and evidence are always taken from the Turkish source; `DomainMatrix::localize()` only replaces these six fields, at read time, with the text in the request language, and a missing file, cell or field falls back to Turkish. `evidence.note` and the reference list count as researcher text and are not translated (ADR-046 §8). The motif and temperament heading also arrives in the request language (`database/data/<lang>/motifs.json`).

7.6.2 Contract tests

The corpus is data; its checking is code. `DomainMatrix::errors()` checks the following rules, and `DomainMatrixTest::test_korpus_sozlesmeye_uyar` requires it to return an empty list (`Backend/tests/Feature/DomainMatrixTest.php`):

1. **The core is complete.** In every branch there must be a cell whose `facet` value is `null` for each of five domains × two bands; a missing combination yields the error "çekirdek hücre eksik" (core cell missing). There must be at least ten cells in every branch.
2. **Identity and consistency.** `id` is unique; `domain` matches the file name and the first letter of the `facet` code matches `trait`; `band` may only be `low` or `high`.
3. **The source is in the reference list.** `evidence.sources` may not be empty and every key must be present in `sources.json`; the grade is one of three values.
4. **Required texts.** `behavior`, `narrative` and `goal_light` may not be empty.
5. **Coefficient check.** If a coefficient pattern of the kind `r = ,21`, `d ≈ .65`, `OR 1,22` (`\b(r|d|OR|HR|RR|β)\s*[=≈]\s*…`) is caught in the six user text fields, an error is raised; coefficients live only in `evidence.note` (ADR-028).
6. **Absolute language.** The second test (`test_hucre_metinleri_turkce_ve_deterministik_degil`) forbids the words "asla" (never), "her zaman" (always), "kesinlikle" (definitely) and "mutlaka" (without fail) in the six user texts.

The translation layer has its own checks: `test_motif_dil_dosyalari_turkceyle_esit` requires the motif files of the four languages to carry the same keys and array lengths as the Turkish source, and `test_persona_istek_dilinde_bilinmeyen_dil_ingilizce` requires the heading to arrive in the request language and an unknown language to fall back to English; the ADR-046 case in `InvariantsTest` verifies that the six user fields of a translated cell are identical to the language file while the `evidence` field remains identical to the Turkish source.

These tests carry the same philosophy as `PersonalityRuleCatalogTest` in Chapter 6: as the evidence changes the record changes, the engine does not; and if the source of a cell is lost, the test breaks.

7.6.3 Barnum check and language rules

The rules set for the language of the report (ADR-028; ADR-031) were carried over to the corpus. The texts are in the second person singular, in the "tendency" mode and linked to the band; no band is given a sentence of the kind "your relationship will break down", "more religious" or "less spiritual"; where there is evidence, every band has a strength; expressions of the "proven" kind are not used (report §1.5). The Barnum risk takes a different form here from the one in Chapter 5: not a sentence that fits everyone, but the narrative of one band also fitting the opposite band. The design decision that reduces this is that every narrative gives "the other side of the coin" in its third sentence and the limit of the tendency in its fourth; in the `finance.C.low` example (§7.8.2) the narrative states together the cost of the low band and the fact that the same band is the one that responds best to intervention. A strength for which there is no evidence is not invented: the `strength` field of the `finance.C.low` cell is `null`, because the notes reported no marked financial advantage in this band.

7.6.4 The "moderate and above" filter and the quality gate

There are 14 `zayıf` cells in the corpus: `finance.O.low`, `finance.A.low`, `finance.C5.low`, `finance.C6.low`, `health.O.high`, `health.O.low`, `relationships.O.high`, `relationships.O.low`, `relationships.E1.high`, `career.E.low`, `spiritual.C.low`, `spiritual.A.low`, `spiritual.E.low`, `spiritual.O6.high` (report §6). Most are the reverse reading of a null finding: deriving a sentence for "low Openness" from the finding "high Openness is unrelated to relationship satisfaction" does not produce evidence. These cells are not rendered on the card; the threshold `config('domain_matrix.min_grade') = 'orta'` is applied inside `forUser()`, and the test `test_zayif_kanitli_hucre_kartta_gorunmez` verifies that no weak cell is returned in any of the five branches. The weak cells remain in the corpus, because, as the "Five branches" section of the report uses the same filter too, they are not visible on any surface today; their existence serves to ensure that the information "no distinguishing finding in this band" is not lost and that they can be updated directly in the second research round.

The quality gate is the same as in ADR-033 §3: if the profile quality is not `ok`, no cell is returned (`test_kalite_uninterpretable_hucre_dondurmez`); at the engine gate, Pusula's `ok` and single-flag `note` levels count as `ok` and the `review` level as `caution` (`PersonalitySignals::qualityLevel()`). For a user with no Pusula profile the endpoint returns an empty list and `quality: none`, and a link to the test is shown in place of the card. The endpoint returns at most four cells per branch (`max_cells`); the "Five branches" section of the report shows all four, while the card in the branch pane, so as to remain scroll-free, shows the first three (the first two in the narrow layout) (`LifeTreeWidget.tsx`). The ordering considers facet cells first (more specific), then the evidence grade.

7.7 Transfer

7.7.1 The branch pane of the Life Tree

ADR-037 §3 chooses the branch pane of the Life Tree as the first home of the matrix: users see their profile while setting a goal in that branch. The pane consists of two cards; on the left the branch (ring, goal rows, new goal), on the right `DomainInsightCard`. The card has three tabs: **You** (motif heading and cell narratives), **Strengths · Cautions** (two columns: strengths and "cautions · areas to grow"), **When setting goals** (goal-setting lights). Each row says by a label which band(s) it comes from ("Conscientiousness · low", "Impulse control · low"); at the top right of the card stands the weakest evidence grade ("evidence moderate+"), and below it the source keys (`DomainInsightCard.tsx`). The data come from the endpoint `GET /api/v1/personality/domain-insights?domain=` (`PersonalityDomainInsightController`); the card sends a request only for a signed-in user and shows a fixed example in the preview (for the career branch only). The endpoint returns up to four cells; so as to remain scroll-free, the card renders the first three in the wide layout and the first two in the narrow layout. The cell texts and the motif heading arrive in the request language (ADR-046 §8; §7.6.1).

7.7.2 No link to action

The third tab of the card carries a decision: the cell does not say which goal to set. The description of the tab states this to the user: "Which goal you set depends on you and your life; here is only *how* setting a goal on this branch tends to hold better for you." ADR-037 records this as "no link to action — which goal is up to the user". The rationale is twofold. The first is the second condition of the addendum to ADR-033: the record does not change a value or remove an option; the "light" adds a text and does not fill in the goal builder. The second is the form of the evidence: the literature says that implementation intentions and breaking work into sub-steps work at low Conscientiousness (Gollwitzer & Sheeran, 2006; Locke & Latham, 2002); it does not say which goal should be set. The light of the `career.C.low` cell shows this distinction: "A small number of goals, each with a visible step this week; the language of 'Tuesday morning, at the desk, 20 minutes' rather than 'by the end of the quarter'. Make the key result something that fills in by itself". The subject of the goal is empty; its form is evidenced.

7.7.3 The "Five branches" section of the report

The same corpus appears in the full report as the "Five branches" section (`FiveDomains.tsx`); it is one of the fourteen sections of the report and is locked in the summary (ADR-045; ADR-081). For the five branches the section gives the cells selected by the user's bands in collapsible blocks: band label ("Conscientiousness · low"), narrative, strength and caution cards, and the goal-setting light. In the first version there were, under each block, a source line that showed the evidence grade label together with `evidence.note`, at the head of the section an introduction saying that the motif is only a name, and at its end a four-sentence "Method" box; these were removed from the report on 22 November 2026 by user decision (header note of `FiveDomains.tsx`). Today the coefficients and the evidence note are not readable in any interface; they stay in the corpus and in the endpoint response. The only place where method information remains in the report is the line in the technical appendix: the findings are sourced cells tied to domain and facet bands; only moderate and strong evidence is shown; motif and temperament do not select cells (`TechnicalNote.tsx`).

In the summary report the section behaves differently from the other locked sections. ADR-081 made the locked sections expandable and placed in each a real opening sentence from the person's own result (`PersonalityReport::summary`, `openings`); "Five branches" is not in this list, because the cells come from an endpoint that requires membership. On the locked "Five branches" page only the introductory sentence of the section and the blurred filler built from the introductions of the other sections are visible; no cell text is sent to the summary (ADR-081 §7). On the server side, `PersonalityTestService::fiveDomains()` calls `DomainMatrix::forUser()` for each branch and adds the result to the `five_domains` field of the report.

7.7.4 Score→band fallback for pre-v4 profiles

Profiles from before form v4 have no `profile.domains` and no percentiles; there is only `personality.traits[].score` (0–100). In this case `PersonalitySignals` bands the score as though it were a percentile (below 40 `low`, above 60 `high`) and derives the domain bands from this; facets and motif are not produced (`Back-end/app/Services/PersonalityRules/PersonalitySignals.php`). The test `test_eski_bicim_profil_puandan_bantlanir` verifies that, for a record carrying only a raw score, the `career.C.high` cell is returned and the `persona` field remains `null`. For current profiles the order is as follows: if the result has a percentile, the band derives from it (below 40 `low`, above 60 `high`); if there is no percentile, the band that Pusula gives from the raw score is used (a score of 40 and below `low`, 60 and above `high`). The source of the percentile changed with ADR-082: instead of the reference group formed by those who completed the test in Rehnüma, the published open data of the IPIP-NEO-120 are used; the `PersonalitySignals` code has not changed, only the source of the `norms` field it reads. The legacy-format fallback is a compromise: a raw score is not a norm-referenced percentile, and this is a relaxed form of the rule of ADR-033 §3, "if there is no percentile, Pusula's band"; it will be removed as pre-v4 records run out (inference).

7.8 Working through sample profiles

7.8.1 The ISTP profile in the sample report

The sample report (`Front-end/src/components/test/report/ornek-rapor.json`) carries a profile in the low band in all five of the five domains: O 30.2; C 28.1; E 13.5; A 27.1; S 34.4 (0–100 scores, all `low`); the mind motif is ISTP. The `norms` section of the file was recomputed from the published norm with ADR-082: the domain percentiles are 1 in O, C, E and A and 12 in S; the facet percentiles are below 40, the exceptions being S3 Mood balance (60) and C6 Deliberation (46); no percentile is given for O6 Flexibility in values. This profile scans the low-band side of the corpus almost completely. Table 7.8 gives the selection in the `five_domains` field of the report.

Table 7.8. Cells selected in the report for the sample ISTP profile.

Branch	Selected cells (order: facet first, then grade)	Eliminated cells and the reason
Career	`career.C.low`, `career.S.low`, `career.O.low`, `career.A.low`	`career.E.low` (weak)
Finance	`finance.S5.low`, `finance.C.low`, `finance.S.low`, `finance.E.low`	`finance.C5.low`, `finance.C6.low`, `finance.O.low`, `finance.A.low` (weak)

Branch	Selected cells (order: facet first, then grade)	Eliminated cells and the reason
Health	`health.C5.low`, `health.S5.low`, `health.C.low`, `health.S.low`	`health.E.low`, `health.A.low` (moderate/strong, but `max_cells` = 4 was full); `health.C6.low` (facet band `mid`, not selected); `health.O.low` (weak)
Relationships	`relationships.S.low`, `relationships.E.low`, `relationships.A.low`, `relationships.C.low`	`relationships.S2.low` (facet band `mid`, not selected); `relationships.O.low`, `relationships.E1.high` (weak)
Spiritual life	`spiritual.S1.low`, `spiritual.S.low`, `spiritual.O.low`	`spiritual.C.low`, `spiritual.E.low`, `spiritual.A.low` (weak)

Note. Source: `ornek-rapor.json`, `five_domains`. In health the `max_cells` limit left out the `health.E.low` cell, which is graded `güçlü`; this is a cost of facet priority and is discussed in §7.9.4. The `five_domains` field in the sample file was produced on 25 October 2026 by `DomainMatrix` selection from the domain and facet bands the file had on that day (`personality.facets[].band`); cells whose facet band was `mid` (e.g. `health.C6.low`, `relationships.S2.low`) were not selected. The field has not been regenerated: in the post-ADR-082 `norms` section of the file the percentile of S2 Managing anger is 39, and since `PersonalitySignals` prefers the percentile to the band, if the same file were passed through the endpoint today `relationships.S2.low` would be the first cell of the relationships branch and `relationships.C.low` would fall outside the four (inference from the code; the selection for the other four branches does not change).

Two features of the selection stand out. First, in three of the five branches (finance, health, spiritual life) the first cell is a facet; this means that the card starts the "your band" sentence from a place more specific than the domain in general. Second, only three cells were returned in spiritual life: because the profile's low C, E and A bands are graded `zayıf` in this branch, they do not appear on the card. That nothing is said to the user in this branch for low Extraversion and low Agreeableness is how the "no empty band" rule looks from the user's side: the cell exists, the sentence does not.

7.8.2 The full text of one cell: `finance.C.low`

The `finance.C.low` cell (grade `orta`), in second place in the finance branch, shows the three text layers of the corpus together.

Narrative (You tab): "What sets you apart in finance is that the distance between knowing and doing is wider for you: time spent on money management and financial strain go with low Conscientiousness, and literacy alone does not close that gap. This shows up as the decision 'I should save' not reaching the end of the month, the statement being opened when it arrives, and subscriptions being forgotten. The other side, and the real news, is this: the approach that works best in this band is not growing willpower but deciding once and letting the system run. The tendency can stay; its outcomes change."

Caution: "Financial literacy training on its own does not turn into savings for you; the knowledge is there, the bridge to practice stays missing." *Strength:* `null`.

Goal-setting light: "Tie the goal to a mechanism that fills by itself: not 'save 2,000 a month' but 'automatic 2,000 transfer on payday is set up'. Let the key result be 'how many months the automatic transfer ran without a break', not 'how much I saved'; let reminders name a specific goal and amount, not a general one."

Evidence note (in the corpus and in the endpoint response; not shown in the interface): "Save More Tomorrow: saving among participants 3.5% → 13.6% (participant/non-participant comparison, low causal grade). Karlan2016, three bank field experiments: the reminder works when it mentions a specific expense;

the frame makes no difference. Letkiewicz2014: the literacy effect appears together with C. No marked financial advantage was reported in this band." Sources: Donnelly et al. (2012), Xu et al. (2015), Letkiewicz and Fox (2014), Thaler and Benartzi (2004), Karlan et al. (2016).

The grade of the cell is `orta` because the behavioural finding rests on a single large panel and a few consistent studies, while the intervention evidence comes from a comparison of low causal grade and one field experiment. The suggestion of the light, "make the key result the number of uninterrupted months of the transfer", is consistent with the "progress framing" rule of Chapter 6 and with the principle of ADR-034 that "light is progress"; this translation, however, is an inference of the product and not of the note, and it is counted in §7.9.2.

7.8.3 The INFJ profile in the demo account

A fully formed sample profile has been loaded into the demo account (INFJ · Deep Counsellor · NF · morning type); the motif texts come from `Back-end/database/data/motifs.json` ("Quiet, intuitive and in pursuit of meaning"). The band data of the profile are in the database, not in the repository; for this reason the cell selection is derived here not from the code but from the mapping (inference). By the mapping of McCrae and Costa (1989), the letters I–N–F–J correspond to the low E, high O, high A and high C bands; the S axis cannot be read from the letters. With these bands the cells the corpus would select are the following: in career `career.C.high`, `career.O.high`, `career.A.high` (and, if available at facet level, `career.C4.high`, `career.O5.high`); in finance `finance.C.high`, `finance.O.high`, `finance.A.high`, `finance.E.low`; in health `health.C.high`, `health.A.high`, `health.E.low`; in relationships `relationships.A.high`, `relationships.C.high`, `relationships.E.low`; in spiritual life `spiritual.O.high`, `spiritual.A.high`, `spiritual.C.high` and, at facet level, rows such as `spiritual.O5.high`, `spiritual.A3.high`, `spiritual.A6.high`, `spiritual.C4.high`. The instructive value of the demo profile is the reverse of the ISTP example: the `spiritual.O.high` cell is the only `güçlü` high-Openness spiritual life row of the corpus, and its narrative opens with "searching is a journey for you, not a loss"; the goal light asks about the kind of learning (exploring, deepening, finishing), and the evidence note states that this question is an inference. One caveat is needed: if the profile has completed stage 2 and the relevant facet bands are also high, then, because of facet priority and the `max_cells` = 4 limit, the four places in the spiritual life branch fill with facet cells graded `orta`, and `spiritual.O.high` and `spiritual.A.high`, graded `güçlü`, fall outside the list (inference from the code; §7.9.4). The name "Deep Counsellor" in the heading of the card enters this selection nowhere; the card of an ESFJ with the same bands would open with a different heading and the same cells (inference).

7.9 Limits and open work

7.9.1 Weak cells are the reverse reading of a null finding

Most of the fourteen `zayıf` cells were derived from the absence of a positive finding, and this is not evidence. The corpus stores them but does not show them; yet storing them is itself a risk: if `min_grade` is lowered in future, or if another surface reads the corpus without the filter, the user sees an

unevidenced sentence of the kind "low Openness is like this in relationships". Against this risk the contract test checks not the value of `min_grade` but only the behaviour of the endpoint; it is recommended that lowering `min_grade` to `zayıf` should require a separate ADR (inference).

7.9.2 Goal lights labelled "(inference)"

The `goal_light` field is the least evidenced layer of the corpus. The notes give evidence on behaviour and intervention; they do not give the *form* of the goal; most of the suggestions about form are the product's translation and are marked in the evidence notes as "(hedef kurma ışığı çıkarım)" (goal-setting light is an inference). In four cells this label occurs directly in the `goal_light` text (`health.O.high`, `health.O.low`, `relationships.E1.high`, `spiritual.E1.high`); in the evidence note of 71 of the 76 cells it is stated that the light is an inference (the phrase "hedef kurma ışığı çıkarım" was counted in the `evidence.note` fields). In other words, the behaviour layer of the corpus is sourced, whereas the goal-form layer is almost entirely inference; the card says this in part through the description of the "When setting goals" tab, but the cell-level label, since the source line of the report has been removed (§7.7.3), is not visible in any interface today. ADR-037 records the verification of these lights with user data as open work.

7.9.3 Turkish norms and culture

In none of the five notes is there a systematic finding for Big Five × branch behaviour in a Turkish-speaking sample. The spiritual life note searches Turkish sources separately and finds no Turkish article giving a direct "Five-Factor × religiosity scale" correlation table; the Turkish findings for death anxiety are conflicting; there is no Muslim or Turkish-speaking sample for the MBTI/Keirsey. The relationships note states that studies in Turkish samples exist but that most are small and cross-sectional and that they were not searched systematically in this session. The loss effect of the colour red (Bazley et al., 2021) depends on culture. The bands themselves are not computed against a sample from Turkey either. With ADR-082 the personality percentiles come not from the reference group of those who completed the test in Rehnüma (ADR-032) but from the published open data of the IPIP-NEO-120 (619,150 people; a separate table for the 60 core items of stage 1 and for the 120 items of stage 2, the mid-rank percentile of every possible total score). ADR-082 records the limits of this: the norm comes from volunteers who completed the English form online, and carrying it over to the Turkish adaptation is an assumption; the sample includes all ages and countries together, and no separate norms by age and sex are used; since the international means are above 50 (53–67 in the domains), the same raw score may fall to a lower percentile than against the previous group; this is expected to make the `low` band more frequent (inference). `ReferenceGroup` remains in the code but is not connected to the report; it can be reused if a norm for Turkey is collected in future. In either case, the literature on which the cell contents rest comes largely from Western samples.

7.9.4 The stage 2 requirement for facets

The twenty-six facet cells can be selected only for profiles in which the stage 2 (deepening) measurement has been completed; the summary profile has only domain bands (ADR-025 §6; ADR-033 §3). The "facet first" ordering of the card therefore never comes into play for a stage 1 user, and the content of the card changes when the same user finishes stage 2. This change is not explained to the user today; a notification of the kind "the card was refreshed after deepening" is open work (inference).

Facet bands also derive from the percentile (§7.7.4); the only exception is O6 Flexibility in values: since the three rewritten O6 items are not present in the published data, no percentile is given for this facet (ADR-082 §4) and its band falls back to Pusula's raw-score band. Since the only O6 cell in the corpus (`spiritual.O6.high`) is graded `zayıf`, this difference is not reflected on any surface today. In addition, when the `max_cells` = 4 limit is combined with the facet-priority ordering, it can leave out domain cells graded `güçlü`, as seen in the health branch of the ISTP example; this trade-off between grade and specificity should be reassessed with a user study.

7.9.5 Plan for verification with Rehnüma data

Section 7 of the synthesis report sets up, in measurable form, the plan that ADR-037 §5 requires. (1) **Band × branch behaviour match:** a dashboard criterion is chosen for each cell and the band difference is tracked; in finance, C band × frequency of budget overruns and completion rate of savings goals, S band × frequency of looking at the budget widget and withdrawal after an overrun (the distinction of Olafsson & Pagel, 2017); in health, C × habit *frequency* (Wilson & Dishman, 2015), S × weight fluctuation and deviation in bedtime; in career, C × delay in closing goals; in spiritual life, C × sustaining a contemplation/worship habit, O × number of learning goals and completion rate; in relationships, E × the habit of one-to-one meetings. The expected effect is small ($r \approx .10$–$.25$), and the sample is planned accordingly. (2) **Facet hypotheses:** low S5 × overruns in the "Wants" category; high C4 × exercise frequency; low C5 × exercise *unrelated* (null hypothesis; $r = .07$ should be replicated); high O5 × learning goals in spiritual life. (3) **Moderation:** whether the effect of the implementation intention field is larger at low/moderate C than at high C (the conflict between Webb et al., 2007 and Elliott et al., 2024); whether the effect of a risk message lies in the combination of low S + low C. (4) **The card itself:** using the events (`widget_view`, `copy_variant_shown`, `rule_overridden`) in the measurement plan referred to by ADR-033 (Chapter 6, Table 6.10), the rate of setting a goal in that branch and sustaining it for 30 days among those who see the card, against those who do not; a personalisation RCT does not exist in the literature, and this may be Rehnüma's first contribution. (5) **Stability:** the retest stability of the Pusula bands (ADR-029) is tracked together with the branch criteria; for a user whose band changes the cell changes too, and this is reported separately. (6) **Record:** the results are added to this chapter, the Rehnüma finding is written into the `evidence.note` field of each cell, and the grade is lowered if necessary. The preconditions of the plan are the same as those of Chapter 6: there is no event logging today (ADR-033, "Preconditions").

7.9.6 The second research round

ADR-037 envisages the second round in order to multiply the facet cells in the relationships and spiritual life branches. The present distribution is unbalanced: there are two facet cells in relationships and ten in spiritual life, and the ten in spiritual life come from a single study. The relationships note records that there is no direct evidence linking A1 Trust, A4 Compliance and E1 Warmth to a relationship outcome; the spiritual life note records that there is no facet-level evidence at the level of spirituality for O3 Openness to feelings and O6 Flexibility in values. In finance there is no direct study with the 30 NEO facets, and the UPPS bridge is conceptual. Three priorities are proposed for the second round (inference): a targeted search for relationship and spirituality studies conducted at facet level; a re-reading of the career coefficients from the primary texts in the light of the correction of Sackett et al. (2022); and a search of DergiPark and YÖK Tez in the Turkish literature with the keys "beş faktör" (five factor) + religiosity/relationship satisfaction/financial behaviour.

Chapter references

Alarcon, G., Eschleman, K. R., & Bowling, N. A. (2009). Relationships between personality variables and burnout: A meta-analysis. *Work & Stress, 23*(3), 244–263. https://doi.org/10.1080/02678370903282600

Alderotti, G., Rapallini, C., & Traverso, S. (2023). The Big Five personality traits and earnings: A meta-analysis. *Journal of Economic Psychology, 94*, 102570. https://doi.org/10.1016/j.joep.2022.102570

Algoe, S. B., Gable, S. L., & Maisel, N. C. (2010). It's the little things: Everyday gratitude as a booster shot for romantic relationships. *Personal Relationships, 17*(2), 217–233. https://doi.org/10.1111/j.1475-6811.2010.01273.x

Ames, D. R., & Flynn, F. J. (2007). What breaks a leader: The curvilinear relation between assertiveness and leadership. *Journal of Personality and Social Psychology, 92*(2), 307–324. https://doi.org/10.1037/0022-3514.92.2.307

Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology, 74*(6), 1531–1544. https://doi.org/10.1037/0022-3514.74.6.1531

Barańczuk, U. (2019). The Five Factor Model of personality and social support: A meta-analysis. *Journal of Research in Personality, 81*, 38–46. https://doi.org/10.1016/j.jrp.2019.05.002

Baranski, E., et al. (2024). Personality and conceptions of religiosity across the world's religions. *Journal of Research in Personality, 110*, 104496. https://doi.org/10.1016/j.jrp.2024.104496

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium. *International Journal of Selection & Assessment, 9*(1–2), 9–30. https://doi.org/10.1111/1468-2389.00160

Bazley, W. J., Cronqvist, H., & Mormann, M. (2021). Visual finance: The pervasive effects of red on investor behavior. *Management Science, 67*(9), 5616–5641. https://papers.ssrn.com/ab_id=2992812

Berry, D. S., & Hansen, J. S. (2000). Personality, nonverbal behavior, and interaction quality in female dyads. *Personality and Social Psychology Bulletin, 26*(3), 278–292. https://doi.org/10.1177/0146167200265002

Bess, T. L., & Harvey, R. J. (2002). Bimodal score distributions and the MBTI: Fact or artifact? *Journal of Personality Assessment, 78*(1), 176–186. https://doi.org/10.1207/S15327752JPA7801_11

Billieux, J., Rochat, L., Rebetez, M. M. L., & Van der Linden, M. (2008). Are all facets of impulsivity related to self-reported compulsive buying behavior? *Personality and Individual Differences, 44*(6), 1432–1442. https://doi.org/10.1016/j.paid.2007.12.011

Bloch, L., Haase, C. M., & Levenson, R. W. (2014). Emotion regulation predicts marital satisfaction: More than a wives' tale. *Emotion, 14*(1), 130–144. https://doi.org/10.1037/a0034272

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin, 130*(6), 887–919. https://doi.org/10.1037/0033-2909.130.6.887

Brown, S., & Taylor, K. (2014). Household finances and the "Big Five" personality traits. *Journal of Economic Psychology, 45*, 197–212. https://eprints.whiterose.ac.uk/82777/3/WRRO_82777.pdf

Buecker, S., Maes, M., Denissen, J. J. A., & Luhmann, M. (2020). Loneliness and the Big Five personality traits: A meta-analysis. *European Journal of Personality, 34*(1), 8–28. https://doi.org/10.1002/per.2229

Buss, D. M. (1991). Conflict in married couples: Personality predictors of anger and upset. *Journal of Personality, 59*(4), 663–688. https://doi.org/10.1111/j.1467-6494.1991.tb00926.x

Carlo, G., Okun, M. A., Knight, G. P., & de Guzman, M. R. T. (2005). The interplay of traits and motives on volunteering. *Personality and Individual Differences, 38*(6), 1293–1305. https://digitalcommons.unl.edu/psychfacpub/9

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation. *Journal of Applied Psychology, 85*(5), 678–707. https://doi.org/10.1037/0021-9010.85.5.678

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Donnelly, G., Iyer, R., & Howell, R. T. (2012). The Big Five personality traits, material values, and financial well-being of self-described money managers. *Journal of Economic Psychology, 33*(6), 1129–1142. https://ideas.repec.org/a/eee/joepsy/v33y2012i6p1129-1142.html

Driver, J. L., & Gottman, J. M. (2004). Daily marital interactions and positive affect during marital conflict among newlywed couples. *Family Process, 43*(3), 301–314. https://doi.org/10.1111/j.1545-5300.2004.00024.x

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance. *Journal of Applied Psychology, 91*(1), 40–57. https://doi.org/10.1037/0021-9010.91.1.40

Duggan, K. A., Friedman, H. S., McDevitt, E. A., & Mednick, S. C. (2014). Personality and healthy sleep: The importance of conscientiousness and neuroticism. *PLoS ONE, 9*(3), e90628. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3961248/

Dyrenforth, P. S., Kashy, D. A., Donnellan, M. B., & Lucas, R. E. (2010). Predicting relationship and life satisfaction from personality in nationally representative samples from three countries. *Journal of Personality and Social Psychology, 99*(4), 690–702. https://pubmed.ncbi.nlm.nih.gov/20718544/

Elliott, M. A., McGroarty, A., & Robertson, D. J. (2024). Using a volitional help sheet to increase university students' attendance at synchronous online lectures: A randomized controlled trial. *British Journal of Educational Psychology, 94*(2), 328–345. https://doi.org/10.1111/bjep.12647

Entringer, T. M., Gebauer, J. E., & Kröger, H. (2023). Big Five personality and religiosity: Bidirectional cross-lagged effects and their moderation by culture. *Journal of Personality, 91*(3), 736–752. https://doi.org/10.1111/jopy.12770

Epton, T., Currie, S., & Armitage, C. J. (2017). Unique effects of setting goals on behavior change: Systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology, 85*(12), 1182–1198. https://doi.org/10.1037/ccp0000260

Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin, 136*(5), 894–914. https://doi.org/10.1037/a0019993

Furnham, A. (1996). The big five versus the big four: The relationship between the MBTI and NEO-PI five factor model. *Personality and Individual Differences, 21*(2), 303–307. https://doi.org/10.1016/0191-8869(96)00033-5

Gardner, W. L., & Martinko, M. J. (1996). Using the MBTI to study managers: A literature review and research agenda. *Journal of Management, 22*(1), 45–83. https://doi.org/10.1177/014920639602200103

Gerlach, G., Herpertz, S., & Loeber, S. (2015). Personality traits and obesity: A systematic review. *Obesity Reviews, 16*(1), 32–63. https://doi.org/10.1111/obr.12235

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology, 38*, 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1

Gottman, J. M., & Levenson, R. W. (2000). The timing of divorce: Predicting when a couple will divorce over a 14-year period. *Journal of Marriage and Family, 62*(3), 737–745. https://doi.org/10.1111/j.1741-3737.2000.00737.x

Goyal, M., et al. (2014). Meditation programs for psychological stress and well-being: A systematic review and meta-analysis. *JAMA Internal Medicine, 174*(3), 357–368. https://www.uned.es/universidad/inicio/dam/jcr:5a55d4bf-c8bf-4797-aca6-c0e85802188d/Goyal%20(2014)%20Meta-analisis%20mindfulness.pdf

Graham, E. K., et al. (2017). Personality predicts mortality risk: An integrative data analysis of 15 international longitudinal studies. *Journal of Research in Personality, 70*, 174–186. https://stacks.cdc.gov/view/cdc/53475/cdc_53475_DS1.pdf

Grant, A. M. (2013). Rethinking the extraverted sales ideal: The ambivert advantage. *Psychological Science, 24*(6), 1024–1030. https://doi.org/10.1177/0956797612463706

Hakulinen, C., et al. (2015). Personality and smoking: Individual-participant meta-analysis of nine cohort studies. *Addiction, 110*(11), 1844–1852. https://pmc.ncbi.nlm.nih.gov/articles/PMC4609271

Hammond, M. M., Neff, N. L., Farr, J. L., Schwall, A. R., & Zhao, X. (2011). Predictors of individual-level innovation at work: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts, 5*(1), 90–105. https://doi.org/10.1037/a0018556

Harris, K., & Vazire, S. (2016). On friendship development and the Big Five personality traits. *Social and Personality Psychology Compass, 10*(11), 647–667. https://doi.org/10.1111/spc3.12287

Heller, D., Watson, D., & Ilies, R. (2004). The role of person versus situation in life satisfaction: A critical examination. *Psychological Bulletin, 130*(4), 574–600. https://doi.org/10.1037/0033-2909.130.4.574

Humbad, M. N., Donnellan, M. B., Iacono, W. G., McGue, M., & Burt, S. A. (2010). Is spousal similarity for personality a matter of convergence or selection? *Personality and Individual Differences, 49*(7), 827–830. https://doi.org/10.1016/j.paid.2010.07.010

Jacques-Hamilton, R., Sun, J., & Smillie, L. D. (2019). Costs and benefits of acting extraverted: A randomized controlled trial. *Journal of Experimental Psychology: General, 148*(9), 1538–1556. https://doi.org/10.1037/xge0000516

Jensen-Campbell, L. A., & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality, 69*(2), 323–362. https://doi.org/10.1111/1467-6494.00148

Jeronimus, B. F., Kotov, R., Riese, H., & Ormel, J. (2016). Neuroticism's prospective association with mental disorders halves after adjustment for baseline symptoms and psychiatric history. *Psychological Medicine, 46*(14), 2883–2906. https://www.cambridge.org/core/journals/psychological-medicine/article/neuroticisms-prospective-association-with-mental-disorders-halves-after-adjustment-for-baseline-symptoms-and-psychiatric-history-but-the-adjusted-association-hardly-decays-with-time-a-metaanalysis-on-59-longitudinalprospective-studies-with-443-313-participants/35DF609ABE724C4F25C247D03C38BB3C

Jokela, M., et al. (2013a). Association of personality with the development and persistence of obesity: A meta-analysis based on individual-participant data. *Obesity Reviews, 14*(4), 315–323. https://doi.org/10.1111/obr.12007

Jokela, M., et al. (2013b). Personality and all-cause mortality: Individual-participant meta-analysis of 3,947 deaths in 76,150 adults. *American Journal of Epidemiology, 178*(5), 667–675. https://pmc.ncbi.nlm.nih.gov/articles/PMC3755650

Joshanloo, M. (2022). Personality trait level and change predict future financial well-being. *Personality and Individual Differences, 191*, 111575. https://flosse.dss.gov.au/flossejspui/handle/10620/18954

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86*(1), 80–92. https://doi.org/10.1037/0021-9010.86.1.80

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*(4), 765–780. https://doi.org/10.1037/0021-9010.87.4.765

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87*(3), 530–541. https://doi.org/10.1037/0021-9010.87.3.530

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*(3), 621–652. https://doi.org/10.1111/j.1744-6570.1999.tb00174.x

Judge, T. A., Livingston, B. A., & Hurst, C. (2012). Do nice guys—and gals—really finish last? The joint effects of sex and agreeableness on income. *Journal of Personality and Social Psychology, 102*(2), 390–407. https://doi.org/10.1037/a0026021

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance. *Journal of Applied Psychology, 98*(6), 875–925. https://doi.org/10.1037/a0033901

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science, 62*(12), 3393–3411. https://doi.org/10.1287/mnsc.2015.2296

Kashdan, T. B., et al. (2018). The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality, 73*, 130–149. https://doi.org/10.1016/j.jrp.2017.11.011

Kern, M. L., & Friedman, H. S. (2008). Do conscientious individuals live longer? A quantitative review. *Health Psychology, 27*(5), 505–512. https://www.gwern.net/doc/psychology/personality/conscientiousness/2008-kern.pdf

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. *Psychological Bulletin, 119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Kong, S. S. (2010). Relationships between Myers-Briggs Type Indicator (MBTI) psychological type and marital satisfaction, divorce proneness, positive affect, and conflict regulation in clinic couples. *Journal of Korean Academy of Nursing, 40*(3), 336–348. https://doi.org/10.4040/jkan.2010.40.3.336

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin, 136*(5), 768–821. https://researchconnect.stonybrook.edu/en/publications/linking-big-personality-traits-to-anxiety-depressive-and-substanc/

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist, 64*(4), 241–256. https://pmc.ncbi.nlm.nih.gov/articles/PMC2792076/

Lambert, N. M., Clark, M. S., Durtschi, J., Fincham, F. D., & Graham, S. M. (2010). Benefits of expressing gratitude: Expressing gratitude to a partner changes one's view of the relationship. *Psychological Science, 21*(4), 574–580. https://doi.org/10.1177/0956797610364003

Lavigne, K. M., Hofman, S., Ring, A. J., Ryder, A. G., & Woodward, T. S. (2013). The personality of meaning in life. *Journal of Positive Psychology, 8*(1), 34–43. https://instituteofcoaching.org/node/34290

Letkiewicz, J. C., & Fox, J. J. (2014). Conscientiousness, financial literacy, and asset accumulation of young adults. *Journal of Consumer Affairs, 48*(2), 274–300. https://doi.org/10.1111/joca.12040

Lipnevich, A. A., Credé, M., Hahn, E., Spinath, F. M., Roberts, R. D., & Preckel, F. (2017). How distinctive are morningness and eveningness from the Big Five factors of personality? A meta-analytic investigation. *Journal of Personality and Social Psychology, 112*(3), 491–509. https://www.ncbi.nlm.nih.gov/pubmed/27977220

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

Malouff, J. M., Thorsteinsson, E. B., Rooke, S. E., & Schutte, N. S. (2007). Alcohol involvement and the Five-Factor Model of personality: A meta-analysis. *Journal of Drug Education, 37*(3), 277–294. https://rune.une.edu.au/web/handle/1959.11/4125

Malouff, J. M., Thorsteinsson, E. B., Schutte, N. S., Bhullar, N., & Rooke, S. E. (2010). The Five-Factor Model of personality and relationship satisfaction of intimate partners: A meta-analysis. *Journal of Research in Personality, 44*(1), 124–127. https://doi.org/10.1016/j.jrp.2009.09.004

Masi, C. M., Chen, H. Y., Hawkley, L. C., & Cacioppo, J. T. (2011). A meta-analysis of interventions to reduce loneliness. *Personality and Social Psychology Review, 15*(3), 219–266. https://doi.org/10.1177/1088868310377394

Matz, S. C., Gladstone, J. J., & Stillwell, D. (2016). Money buys happiness when spending fits our personality. *Psychological Science, 27*(5), 715–725. https://doi.org/10.1177/0956797616635200

McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality, 57*(1), 17–40. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x

McMahan, E. A., & Estes, D. (2015). The effect of contact with natural environments on positive and negative affect: A meta-analysis. *Journal of Positive Psychology, 10*(6), 507–519. https://instituteofcoaching.org/node/35082

Molloy, G. J., O'Carroll, R. E., & Ferguson, E. (2014). Conscientiousness and medication adherence: A meta-analysis. *Annals of Behavioral Medicine, 47*(1), 92–101. https://doi.org/10.1007/s12160-013-9524-4

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships, 25*(6), 889–922. https://doi.org/10.1177/0265407508096700

Mosca, I., & McCrory, C. (2016). Personality and wealth accumulation among older couples: Do dispositional characteristics pay dividends? *Journal of Economic Psychology, 56*, 1–19. https://mural.maynoothuniversity.ie/id/eprint/11388/1/IM_Personality_2016.pdf

Mund, M., & Neyer, F. J. (2014). Treating personality–relationship transactions with respect: Narrow facets, advanced models, and extended time frames. *Journal of Personality and Social Psychology, 107*(2), 352–368. https://doi.org/10.1037/a0036719

Neyer, F. J., & Lehnart, J. (2007). Relationships matter in personality development: Evidence from an 8-year longitudinal study across young adulthood. *Journal of Personality, 75*(3), 535–568. https://doi.org/10.1111/j.1467-6494.2007.00448.x

Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology, 58*(2), 367–408. https://doi.org/10.1111/j.1744-6570.2005.00515.x

Noftle, E. E., & Shaver, P. R. (2006). Attachment dimensions and the Big Five personality traits: Associations and comparative ability to predict relationship quality. *Journal of Research in Personality, 40*(2), 179–208. https://doi.org/10.1016/j.jrp.2004.11.003

Oehler, A., Wendt, S., Wedlich, F., & Horn, M. (2018). Investors' personality influences investment decisions: Experimental evidence on extraversion and neuroticism. *Journal of Behavioral Finance, 19*(1), 30–48. https://doi.org/10.1080/15427560.2017.1366495

Olafsson, A., & Pagel, M. (2017). *The ostrich in us: Selective attention to financial accounts, income, spending, and liquidity* (NBER Working Paper No. 23945). https://www.nber.org/papers/w23945

Oshio, A., Taku, K., Hirano, M., & Saeed, G. (2018). Resilience and Big Five personality traits: A meta-analysis. *Personality and Individual Differences, 127*, 54–60. https://doi.org/10.1016/j.paid.2018.01.048

Parks-Leduc, L., Feldman, G., & Bardi, A. (2015). Personality traits and personal values: A meta-analysis. *Personality and Social Psychology Review, 19*(1), 3–29. https://pure.royalholloway.ac.uk/en/publications/personality-traits-and-personal-values-a-meta-analysis/

Pfattheicher, S., et al. (2017). Old wine in new bottles? The case of self-compassion and neuroticism. *European Journal of Personality, 31*(2), 160–169. https://doi.org/10.1002/per.2097

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research, 57*(3), 210–221. https://doi.org/10.1037/1065-9293.57.3.210

Prinzie, P., Stams, G. J. J. M., Deković, M., Reijntjes, A. H. A., & Belsky, J. (2009). The relations between parents' Big Five personality factors and parenting: A meta-analytic review. *Journal of Personality and Social Psychology, 97*(2), 351–362. https://doi.org/10.1037/a0015823

Rhodes, R. E., & Smith, N. E. I. (2006). Personality correlates of physical activity: A review and meta-analysis. *British Journal of Sports Medicine, 40*, 958–965. https://bjsm.bmj.com/content/40/12/958

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality. *Perspectives on Psychological Science, 2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin, 143*(2), 117–141. https://doi.org/10.1037/bul0000088

Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five personality factors and personal values. *Personality and Social Psychology Bulletin, 28*(6), 789–801. https://journals.sagepub.com/doi/10.1177/0146167202289008

Rose, P., & Segrist, D. J. (2014). Negative and positive urgency may both be risk factors for compulsive buying. *Journal of Behavioral Addictions, 3*(2), 128–132. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4117284/

Ross, C. F. J., & Francis, L. J. (2010). The relationship of intrinsic, extrinsic, and quest religious orientations to Jungian psychological type among churchgoers in England and Wales. *Mental Health, Religion & Culture, 13*(7–8), 805–819. https://doi.org/10.1080/13674670802207462

Rudolph, C. W., Lavigne, K. N., & Zacher, H. (2017). Career adaptability: A meta-analysis. *Journal of Vocational Behavior, 98*, 17–34. https://doi.org/10.1016/j.jvb.2016.09.002

Rustichini, A., DeYoung, C. G., Anderson, J. E., & Burks, S. V. (2016). Toward the integration of personality theory and decision theory in explaining economic behavior: An experimental investigation. *Journal of Behavioral and Experimental Economics, 64*, 122–137. https://doi.org/10.1016/j.socec.2016.04.019 (secondary citation)

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection. *Journal of Applied Psychology, 107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Saroglou, V. (2002). Religion and the five factors of personality: A meta-analytic review. *Personality and Individual Differences, 32*(1), 15–25. https://www.psyreli.org/_files/ugd/ba70ca_9734c4f5bea2409dbf9fe0f286621d16.pdf

Saroglou, V. (2010). Religiousness as a cultural adaptation of basic traits: A Five-Factor Model perspective. *Personality and Social Psychology Review, 14*(1), 108–125. https://research.dial.uclouvain.be/entities/publication/8eb1e87d-7b39-4740-a9b1-f558fca4236d

Saucier, G. (2019). Is religiousness a form of variation in personality, or in culture, or neither? Conceptual issues and empirical indications. *Archive for the Psychology of Religion, 41*(3), 216–223. https://doi.org/10.1177/0084672419894677

Shapiro, G. K., & Burchell, B. J. (2012). Measuring financial anxiety. *Journal of Neuroscience, Psychology, and Economics, 5*(2), 92–103. https://crr.bc.edu/avoidance-comes-with-financial-anxiety/

Sheeran, P., Listrom, O., & Gollwitzer, P. M. (2024). The when and how of planning: Meta-analysis of the scope and components of implementation intentions in 642 tests. *European Review of Social Psychology*. https://doi.org/10.1080/10463283.2024.2334563

Sicherman, N., Loewenstein, G., Seppi, D. J., & Utkus, S. P. (2016). Financial attention. *Review of Financial Studies, 29*(4), 863–897. https://doi.org/10.1093/rfs/hhv073

Sirois, F. M., Kitner, R., & Hirsch, J. K. (2015). Self-compassion, affect, and health-promoting behaviors. *Health Psychology, 34*(6), 661–669. https://pubmed.ncbi.nlm.nih.gov/25243717/

Solomon, B. C., & Jackson, J. J. (2014). Why do personality traits predict divorce? Multiple pathways through satisfaction. *Journal of Personality and Social Psychology, 106*(6), 978–996. https://doi.org/10.1037/a0036190

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240–261. https://doi.org/10.1037/0033-2909.124.2.240

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin, 133*(1), 65–94. https://doi.org/10.1037/0033-2909.133.1.65

Steger, M. F., Kashdan, T. B., Sullivan, B. A., & Lorentz, D. (2008). Understanding the search for meaning in life. *Journal of Personality, 76*(2), 199–228. https://mason.gmu.edu/~tkashdan/publications/searchmeaning_JP.pdf

Stein, R., & Swan, A. B. (2019). Evaluating the validity of Myers-Briggs Type Indicator theory: A teaching tool and window into intuitive psychology. *Social and Personality Psychology Compass, 13*(2), e12434. https://doi.org/10.1111/spc3.12434

Stephan, Y., Sutin, A. R., Bayard, S., Križan, Z., & Terracciano, A. (2018). Personality and sleep quality: Evidence from four prospective studies. *Health Psychology, 37*(3), 271–281. https://pmc.ncbi.nlm.nih.gov/articles/PMC5837948

Stieger, M., Flückiger, C., Rüegger, D., Kowatsch, T., Roberts, B. W., & Allemand, M. (2021). Changing personality traits with the help of a digital personality change intervention. *Proceedings of the National Academy of Sciences, 118*(8), e2017548118. https://doi.org/10.1073/pnas.2017548118

Strickhouser, J. E., Zell, E., & Krizan, Z. (2017). Does personality predict health and well-being? A metasynthesis. *Health Psychology, 36*(8), 797–810. https://europepmc.org/article/MED/28277701

Sutin, A. R., Ferrucci, L., Zonderman, A. B., & Terracciano, A. (2011). Personality and obesity across the adult life span. *Journal of Personality and Social Psychology, 101*(3), 579–592. https://europepmc.org/article/MED/21744974

Sutin, A. R., et al. (2016). The five-factor model of personality and physical inactivity: A meta-analysis of 16 samples. *Journal of Research in Personality, 63*, 22–28. https://doi.org/10.1016/j.jrp.2016.05.001

Swider, B. W., & Zimmerman, R. D. (2010). Born to burnout: A meta-analytic path model of personality, job burnout, and work outcomes. *Journal of Vocational Behavior, 76*(3), 487–506. https://doi.org/10.1016/j.jvb.2010.01.003

Tam, K.-P. (2013). Concepts and measures related to connection to nature: Similarities and differences. *Journal of Environmental Psychology, 34*, 64–78. https://doi.org/10.1016/j.jenvp.2013.01.004

Terracciano, A., & Costa, P. T. (2004). Smoking and the Five-Factor Model of personality. *Addiction, 99*(4), 472–481. https://pmc.ncbi.nlm.nih.gov/articles/PMC2376761/

Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy, 112*(S1), S164–S187. https://doi.org/10.1086/380085

Tsaousis, I. (2010). Circadian preferences and personality traits: A meta-analysis. *European Journal of Personality, 24*(4), 356–373. https://doi.org/10.1002/per.754

Turiano, N. A., et al. (2012). Personality and substance use in midlife: Conscientiousness as a moderator and the effects of trait change. *Journal of Research in Personality, 46*(3), 295–305. https://pmc.ncbi.nlm.nih.gov/articles/PMC3388488

Webb, T. L., Christian, J., & Armitage, C. J. (2007). Helping students turn up for class: Implementation intentions and conscientiousness. *Learning and Individual Differences, 17*(4), 316–327. https://eprints.whiterose.ac.uk/id/eprint/10755

White, M. P., et al. (2019). Spending at least 120 minutes a week in nature is associated with good health and wellbeing. *Scientific Reports, 9*, 7730. https://doi.org/10.1038/s41598-019-44097-3

Whiteside, S. P., & Lynam, D. R. (2001). The Five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*(4), 669–689. https://doi.org/10.1016/S0191-8869(00)00064-7

Wilmot, M. P., & Ones, D. S. (2019). A century of research on conscientiousness at work. *Proceedings of the National Academy of Sciences, 116*(46), 23004–23010. https://doi.org/10.1073/pnas.1909972116

Wilson, K. E., & Dishman, R. K. (2015). Personality and physical activity: A systematic review and meta-analysis. *Personality and Individual Differences, 72*, 230–242. https://doi.org/10.1016/j.paid.2014.08.023

Xu, Y., Beller, A. H., Roberts, B. W., & Brown, J. R. (2015). Personality and young adult financial distress. *Journal of Economic Psychology, 51*, 90–100. https://experts.illinois.edu/en/publications/personality-and-young-adult-financial-distress/

Zelenski, J. M., Santoro, M. S., & Whelan, D. C. (2012). Would introverts be better off if they acted more like extraverts? *Emotion, 12*(2), 290–303. https://doi.org/10.1037/a0025169

Zhao, H., & Seibert, S. E. (2006). The Big Five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology, 91*(2), 259–271. https://doi.org/10.1037/0021-9010.91.2.259

Zhao, H., Seibert, S. E., & Lumpkin, G. T. (2010). The relationship of personality to entrepreneurial intentions and performance: A meta-analytic review. *Journal of Management, 36*(2), 381–404. https://doi.org/10.1177/0149206309335187

Repository documents. ADR-025, ADR-028, ADR-029, ADR-031, ADR-032, ADR-033 (+addendum), ADR-034, ADR-035, ADR-037, ADR-045, ADR-046, ADR-081, ADR-082 (`docs/decisions/`); `docs/research/reports/Beş dal karakter matrisi.md`; `docs/research/research_notes/Beş dal karakter matrisi/` (five notes); `Back-end/database/data/domain_matrix/` (`career.json`, `finance.json`, `health.json`, `relationships.json`, `spiritual.json`, `sources.json`); `Back-end/database/data/motifs.json`; `Back-end/config/domain_matrix.php`; `Back-end/app/Services/PersonalityRules/DomainMatrix.php`; `Back-end/app/Services/PersonalityRules/PersonalitySignals.php`; `Back-end/app/Http/Controllers/Api/V1/PersonalityDomainInsightController.php`; `Back-end/app/Services/PersonalityTestService.php`; `Back-end/tests/Feature/DomainMatrixTest.php`; `Front-end/src/widgets/goals/lifeTree/DomainInsightCard.tsx`; `Front-end/src/components/test/report/pages/FiveDomains.tsx`; `Front-end/src/components/test/report/ornek-rapor.json`; `Front-end/src/widgets/KARAKTER-IZI.md`.

8. Ethics, Privacy and Licences

> **This is not legal advice.** This chapter records how Rehnüma character analysis was designed with respect to consent, the protection of personal data, retention, licences and the commercial model, and which points remain open. No assessment in this chapter replaces the opinion of a legal adviser. The research notes did not examine the KVKK dimension, consent for those under 18, or distance-selling conditions (Rehnüma personality test psychometrics report, "Points requiring legal approval"); every sentence on these subjects has been left to legal review before launch.

A personality test carries the claim of telling a person something they did not know about themselves. This claim gives rise to three kinds of responsibility. The first concerns data: the answers and the profile derived from them are among a person's most intimate data. The second concerns interpretation: an automated report can influence how a person sees themselves and the decisions they make (Chapter 5.1). The third concerns intellectual property: the instruments are the work of others and come with conditions of use. This chapter addresses, in order, consent (8.1), data protection principles and data minimisation (8.2), retention and deletion (8.3), licences (8.4), the limits in the health and finance modules (8.5), the ethical tensions created by the commercial model (8.6) and algorithmic fairness (8.7). Under each heading we first state the decision taken, then its counterpart in the code, and finally the open questions.

8.1 Consent and consent version; parental consent under 18

8.1.1 The consent step

The test requires a consent step before any answers are given (`Front-end/src/components/test/steps/About.tsx`). On the same screen the person enters their life stage (high school, university, working, searching) and year of birth; if they wish, they also choose a target occupation from the occupation catalogue (ADR-074; free text is not accepted, and only the slug of the occupation is stored in the bridge record). The text of the consent box is as follows: "I agree to my answers being processed to produce a personality and interest profile. The answers are sent for scoring to Rehnüma's own test service (Pusula); my name and email are not sent." It is not possible to continue without ticking the box. The opening sentence of the screen says that the person's situation, year of birth and consent are collected; when the year of birth falls below the age of 18, the file card on the screen shows the note "Under 18: a guardian's consent is required."

Consent is checked at three layers. The front end makes the box mandatory. The Rehnüma back end checks the `consent` field with the `required` and `accepted` rules in `Back-end/app/Http/Requests/StartPersonalityTestRequest.php`. Pusula-API repeats the same check in `AssessmentController` and additionally requires the `consent_version` field to equal the version in its own configuration. The Rehnüma client (`Back-end/app/Services/Pusula/PusulaClient.php`) sends this version from the `services.pusula.consent_version` setting; on both sides the current value is

`2026-10`. When opening the test, Pusula-API stores the consent version and the time of consent (`consented_at`) and, for those under 18, also the time of parental consent (`guardian_consented_at`) (`Pusula-API/database/migrations/2026_10_23_000003_create_assessment_tables.php`).

The function of the consent version is to make old and new consents distinguishable when the consent text changes. Because Pusula-API accepts only the current version in its configuration, the version must change when the text changes and the setting of the two services must be updated together; otherwise all new tests are rejected with a 422 error. This tight coupling is deliberate: it prevents new processing from starting on the basis of consent given to an old text (inference). Its cost is that a version change requires a simultaneous release in the two services.

Where the consent record is kept is also a decision. The consent version and time are held in Pusula-API, in the same row as the test itself; Rehnüma's `pusula_links` table has no consent field. In this way the evidence of consent lives together with the data that the consent covers, and is deleted together with it when the test is deleted. However, this means that the consent version under which the data held on the Rehnüma side (e-mail, copy of the result) was collected is not visible in Rehnüma's own records. We recommend that the consent version also be kept in the bridge record for the Rehnüma copy.

8.1.2 The distinction between the privacy notice and explicit consent

The current box combines two things: the error text of the screen names the box "aydınlatma metnini onayla" ("Accept the information notice", that is, the privacy notice), whereas the box itself expresses consent to processing. Turkish data protection legislation treats the obligation to inform and explicit consent as separate concepts; the privacy notice is a provision of information and does not require approval, whereas explicit consent is an approval that relates to a specific matter, is based on information and is declared of one's own free will. A separate privacy notice was written later: ADR-058 rebuilt the privacy, terms of use and KVKK notice pages with claims extracted from the code (`/privacy`, `/terms`, `/data-protection`); until the title and address of the data controller are added, these pages are published with a "draft" banner. However, the consent screen contains no link to this text; the text of the box is a short summary. Linking the consent screen to the privacy notice, and making the consent box relate only to consent, is the first item left to legal review.

8.1.3 Age and parental consent

The year-of-birth field does not accept those younger than 12 or older than 100; parental consent is mandatory for those under 18 (`StartPersonalityTestRequest::withValidator`). Parental consent is a second box: "I am under 18; my parent or guardian consents to my taking this test." The code review revealed three points at this step.

First, **age is calculated from the year of birth alone.** `isMinor` subtracts the year of birth from the current year and treats the person as a minor if the result is less than 18. Because the birthday is not known, someone who has not yet turned 18 within the year may be counted as an adult. For example, on 23 October 2026 a person born in 2008 whose birthday is in December is 17 years old, but the system counts them as 18 and does not ask for parental consent. In terms of its direction, this error is not protective. Asking for the month of birth as well, or also requiring parental consent from those in the boundary year, would close this gap.

Second, **there is a difference in the age limit between the target audience and the contract.** The contract and the interface accept the test from the age of 12; the privacy page also says "The analysis is available from age 12" (ADR-058). The target audience of the research notes, by contrast, is 14 years and over, and item readability was reviewed for the age of 14 (8.7.3). The comprehensibility of the items has not been tested for a user aged 12–13. Either the lower limit must be raised to 14 or a separate decision must be taken for ages 12–13.

Third, **parental consent rests on a declaration.** Whether the person ticking the box is a parent is not verified. The ITC (2005) guidelines accept that in open mode the identity of the test taker cannot be guaranteed and require this to be stated as a limitation. Whether declaration-based parental consent is sufficient for the processing of adolescent data is an open legal question (ADR-026, "Awaiting legal approval"). The components of the consent flow are summarised in Table 8.1.

Table 8.1. Components of the consent flow and their records.

Component	Where it is asked	Where it is checked	Where it is recorded	Open question
Consent to processing	`About.tsx`, mandatory box	`StartPersonalityTestRequest` (`accepted`); Pusula-API `AssessmentController`	Pusula-API `assessments.consented_at`	Separating out the privacy notice
Consent version	Not shown to the user	Pusula-API: current version only (`config/pusula.php`)	Pusula-API `assessments.consent_version`	Keeping it in the Rehnüma bridge record as well
Age	Year-of-birth field	12 ≤ age ≤ 100 (difference in years)	Rehnüma `pusula_links.birth_year`; Pusula-API `birth_year`	Month of birth; a lower limit of 14
Parental consent	Second box under 18	Separately in Rehnüma and in Pusula-API	Pusula-API `guardian_consented_at`	Legal sufficiency of the declaration
Target occupation (optional)	`OccupationField.tsx`, selection from the catalogue	A test is not opened with a slug that is not in the catalogue	Rehnüma `pusula_links.target_occupation` (slug only)	—
E-mail	After the sections, at the account step (`Register.tsx`; ADR-073)	The result is not shown without an account; for an unregistered bridge record the server makes the e-mail mandatory for scoring (`requireEmail`)	Rehnüma `pusula_links.email`; the user record once the account is opened	Consistency of the e-mail purpose with the text (8.3.4)

Note. Source: the code in the repository, 22 November 2026.

8.2 Data protection principles and data minimisation

8.2.1 General principles

Law No. 6698 on the Protection of Personal Data (KVKK) lists general principles for the processing of personal data: lawfulness and fairness; accuracy and, where necessary, being kept up to date; processing for specified, explicit and legitimate purposes; being relevant, limited and proportionate to the purpose for which the data are processed; and retention for the period laid down in the relevant legislation or necessary for the purpose for which the data are processed. These principles were not examined separately in the research notes of this work and are cited here as a framework for organising the design decisions. The ITC (2005) guidelines likewise require test data to be stored and deleted in accordance with local data protection legislation (scoring and norms note, S5).

In Rehnüma's architecture the most prominent application of these principles is the principle of being "limited and proportionate to the purpose", and it was built through the separation between two services.

8.2.2 Service separation: no personal data goes to Pusula-API

ADR-027 forked the core of the test into the Rehnüma repository as a separate Laravel application with a separate database. The consequence of this decision for privacy is written in the second clause of the ADR: no personal data goes to Pusula-API; only the year of birth and the answers do. The note at the head of `PersonalityTestService` repeats the same principle at the level of the code. The request that opens the test carries the year of birth, the consent and parental-consent flags and the consent version (`PusulaClient::startAssessment`); every request also sends the request language (`Accept-Language`, ADR-046). If a target occupation has been chosen, the slug of the occupation is queried together with the test token for the fit calculation (ADR-074). Name, e-mail, life stage and account identifier are not sent; these fields do not exist in Pusula-API's test table. The privacy page states that the life stage also goes to the test service (ADR-058 Decision 1); the code keeps the life stage only in the Rehnüma bridge record, and the page text should be corrected accordingly. Pusula-API has no user accounts. Every request carries the `Authorization: Bearer <PUSULA_SERVICE_KEY>` header, and if the key is not defined the service is closed (`Pusula-API/AGENTS.md`). The identity of a test is the test token, which is a UUID; the link between this token and a person is established only in the `pusula_links` table in Rehnüma.

In the first design of ADR-025, the bridge opened a shadow account with an e-mail address in Pusula for every test and stored the password and session key of this account in encrypted form. ADR-027 removed the shadow account; the three columns concerned were dropped by the migration `2026_10_23_100000_drop_shadow_account_from_pusula_links.php`. The gain is not simplicity alone: it means fewer encrypted secrets, fewer moving parts and no contact information at all in the test service. In the event of a data breach, the Pusula-API database on its own does not say to whom the answers belong (inference: year of birth, answers and timestamps together carry a low but non-zero risk of re-identification).

The bridge itself uses two secrets. The `link_token` held by the client enables an unregistered visitor to return to the test. The test token in Pusula-API (`assessment_token`), by contrast, stays only in Rehnüma and is marked as `Hidden` in the `PusulaLink` model; it is not sent to the client in any re-

sponse (`Back-end/app/Models/PusulaLink.php`). The front end never talks to Pusula-API; all requests pass through the proxy endpoints of the Rehnüma back end (`/api/v1/test/*`) (ADR-025 Decision 2). In this way the address and key of the test service never reach the browser. The data inventory is given in Table 8.2.

Table 8.2. Data processed in character analysis.

Data	Where it is held	Purpose	Retention
Answers and page durations	Pusula-API	Scoring, quality indicators	At most 30 days for a test left without opening an account; with the account for a registered test
Year of birth	Pusula-API and the Rehnüma bridge record	Parental consent, filtering of resource recommendations by age (`ResourceRecommender`), age analysis	Same as the answers
Consent version and times	Pusula-API	Evidence of consent	With the test
Result (scores, profile, matches)	Pusula-API; a copy in the Rehnüma bridge record and in `personality_profiles`	Report; character trace on the dashboard	Produced after the account is opened (ADR-073); with the account
E-mail	Rehnüma only (`pusula_links`, user record)	Account; transactional messages (ADR-064)	With the account; 30 days in a bridge record left without an account being opened
Life stage	Rehnüma bridge record and profile	Stored for the report; student/employee branching has not yet been implemented (note in `About.tsx`)	Same as the answers
Target occupation (slug)	Rehnüma bridge record	Showing fit with the target occupation in the report (ADR-074)	Same as the answers; can be removed from the report
Report feedback	Rehnüma `report_feedback`	Barnum pruning (Chapter 5.8)	Together with the account
Anonymous export	Analysis file (JSONL)	Psychometric analysis	For the duration of the analysis; no policy has been written

Note. Source: the `pusula_links`, `personality_profiles` and `report_feedback` migrations; the Pusula-API `assessments` table; the `ExportResponses` command. The expression "with the account" means deletion when the account is deleted; the deletion chain is detailed in 8.3.

8.2.3 The limit of the anonymity claim

The entrance screen describes the fate of the data as follows: "Your answers are stored without being linked to your identity. If you don't sign up, they are deleted after 30 days." (`entrance/Entrance.tsx`; text in `messages/tr/test.json`). The trust page, for its part, says "Your name, email or account are not sent to the measurement service; the test is scored under a random identifier. The link to your identity is kept only in your Rehnüma account." The sentence on the trust page describes the architecture correctly. The sentence on the entrance screen, however, needs to be narrowed at the level of the system: in Pusula-API the answers are held without being linked to an identity, but in Rehnüma the bridge row holds the e-mail, the test token and the full copy of the result in the same row. For the system as a whole, therefore, the data are not anonymous but pseudonymised. That the sentence a per-

son reads describes the architecture correctly is the basic condition of the trust design (Chapter 5.7); the proposed correction is a service-level statement such as "Your answers are stored in the test service without your name and e-mail".

The same distinction applies to the analysis data. The `pusula:export-responses` command (`Pusula-API/app/Console/Commands/ExportResponses.php`) writes, in place of the test token, the first 12 characters of its SHA-256 digest, together with the age, the timestamp, the items, the raw answers, the page durations, and the quality and scores. The note in the command says "no identity". This is correct, but because the digest is derived from the token, anyone who knows the token can find the row; for someone with access to the Rehnüma database the file is pseudonymised, not anonymous (inference). There is as yet no written policy on where the analysis files will be kept, who will have access to them and when they will be deleted.

8.2.4 Questions of profiling and special-category data

Two questions have been left to legal review. The first is the nature of the personality profile. The legislation counts health data among the special categories of personal data. Personality and interest scores are not in themselves health data; in Rehnüma, however, the same account lives alongside the health modules (measurement, nutrition, medication plan), and the character trace uses personality data to change the language of these modules (Chapter 6). The nature of a combined profile needs to be assessed separately.

The second is profiling. The report is produced entirely automatically, and the character trace changes the language and defaults of some surfaces on the person's dashboard according to personality data. ADR-033 set this up within strict limits: a rule only reduces or softens, does not hide information, does not change a value, can be undone with a single click, and a "Why are we showing it this way?" explanation is envisaged on every surface. These limits aim to prevent profiling from being used against the person, but they do not remove the existence of profiling. The person needs to know this at the consent screen; the current consent text says only "to produce a personality and interest profile" and does not mention the use of the profile on the dashboard. We recommend a separate notice at the registration step.

8.3 Retention and deletion

8.3.1 Unregistered test: 30 days

Tests left without registration, including the e-mail, are deleted after 30 days (ADR-025, closed question 2). The period is set when the bridge record is opened: `PersonalityTestService::start` writes the `expires_at` field of the bridge record by adding `services.pusula.anonymous_ttl_days` (30) days to the current time. Every request made to an expired bridge record receives a 410 response ("This test has expired; you can start again.", `guardAlive`). On registration the `user_id` field of the bridge record is filled and the row becomes exempt from pruning; `isExpired` and the `stale` scope target only bridge records that are not attached to a user (`PusulaLink.php`).

The starting point of the period is a design decision and needs to be documented: the 30 days are counted from the moment the test starts, not from the last activity. Because the result is no longer produced without an account being opened (ADR-073), what is deleted in an unregistered bridge record is mostly the answers: the answers of a person who finishes the sections on day 29 and does not open an account are deleted one day later. This choice is in the direction of data minimisation and matches the period the person is told on the entrance screen, "deleted after 30 days"; the screen does not, however, say from what point the period is counted.

8.3.2 The pruning command and the order of deletion

Deletion is carried out in two services, in a specific order. In Rehnüma the `pusula:prune-links` command runs every day at 04:10, without overlapping (`Back-end/routes/console.php`). The command calls the `PersonalityTestService::pruneStale` method. For every expired unregistered bridge record, the method first deletes the test in Pusula-API through the `DELETE /api/assessments/{token}` endpoint and then deletes the bridge row. If Pusula-API cannot be reached, the row remains and is retried on the next run.

This order is deliberate. The reverse order, that is, deleting the bridge record first and the test afterwards, would, on a night when Pusula-API could not be reached, lead to the loss of the test token and to an orphaned test in Pusula-API that could no longer be deleted. In the current order the token is kept in Rehnüma until its deletion in the test service has been confirmed. Where deletion fails, the data live one day longer, but their trace is never lost.

Pusula-API also has its own safety net (`Pusula-API/routes/console.php`). Every day, tests that are incomplete and have not been updated for more than 30 days are deleted. As the note in the code says, this is for those that escape Rehnüma's deletion endpoint. The scope of the safety net is limited: it targets only tests left half-finished. The deletion of a test that has been completed but not registered depends entirely on Rehnüma's command. If the bridge record of a completed test is deleted by some other route (e.g. manual intervention in the database), an orphaned result may remain in Pusula-API (inference).

Two further small details should be recorded. The first is the time zone: the application time zone of the Rehnüma back end comes from an environment variable and its default is UTC (`Back-end/config/app.php`); the 04:10 of the pruning is in this time zone. The product's decision to use `Europe/Istanbul` in day calculations (ADR-003) does not create a correctness problem for deletion, because the period is stored as an instant (`expires_at`). The second is that the Pusula-API database keeps UTC and uses `Europe/Istanbul` for display (`config/pusula.php`).

8.3.3 The registered user and deletion of the account

The data of a registered user live together with the account. When Version 1.1 was being written, the Rehnüma API had no account deletion endpoint open to the user; ADR-067 closed this gap. Account deletion is immediate and permanent (`DELETE /account`; `AccountSecurity::delete`): an account with a password confirms by typing its password, and an account that signs in only with Google by typing its e-mail address; there is no waiting period and no undo. The order of deletion is the same as in the pruning command: for every bridge record attached to the user, the test in Pusula-API is deleted first, then the bridge row. If Pusula-API cannot be reached, the `expires_at` value of the bridge record is

pulled to that moment; when the user row is deleted the `user_id` field is emptied by `nullOnDelete`, the bridge record falls within the `stale` scope and the nightly `pusula:prune-links` deletes on both sides. The `personality_profiles` and `report_feedback` rows go together with the user through `cascadeOnDelete`. Order records are not deleted, for reasons of statutory retention; `orders.user_id` is emptied and the record loses its link to the account (ADR-067 Decision 3). An administrator account cannot be deleted through this endpoint.

Account deletion is tested in `AccountSecurityTest` (`test_deleting_the_account_needs_confirmation_and_removes_everything`); because this test does not set up an account that has a bridge record, it does not verify the deletion branch in the test service or the fallback when the service is unreachable (inference: this branch was read from the code).

8.3.4 Deletion whenever the user wishes

In Version 1.1 the trust page said "You can delete the test whenever you like", and this sentence was not matched by the code. The page text has since been corrected: today it says "Tests left without signing up are deleted automatically after 30 days. If you want yours deleted sooner, just write to noreply@rehnuma.co; an in-app delete button is planned." (`Front-end/src/app/[locale]/(marketing)/analysis/trust/page.tsx`; text in `messages/tr/trust.json`). Pusula-API provides the deletion endpoint (`DELETE /api/assessments/{token}`), and in the Rehnüma client this endpoint is defined by the `deleteAssessment` method; the method is called from the pruning command and from account deletion. In Rehnüma's test routes there is still no endpoint through which the user can delete only their test, and no button in the interface: a registered user can have their test deleted only by deleting their account, and an unregistered user by requesting it by e-mail. The page now says this as it is; the trust design (Chapter 5.7) rests precisely on what is written on the page being true. That the request address is a mailbox that is answered, and how requests will be met, should be tied to a written process.

The gap on the e-mail side has also been closed, but a new gap has taken its place. The e-mail screen in Version 1.1 promised to send the link to the result, whereas the code did not send it. ADR-064 introduced transactional messages: when the test is scored, the "your analysis is ready" message is sent (`PersonalityTestService::complete`), and with ADR-073 the e-mail screen was removed and the address came to be collected at the account step. The trust page, on the other hand, still says "Your email is used only so your result is not lost; no marketing emails are sent." The address is now the address of the account, and transactional messages relating to the account (verification link, password reset, new-device alert, order notification) also go to this address; the privacy page lists these, whereas the word "only" on the trust page describes the purpose more narrowly than it is and should be updated. The deletion paths and their status are summarised in Table 8.3.

Table 8.3. Deletion paths and their status.

Path	Trigger	What is deleted	Code	Status
Expiry of an unregistered bridge record	Daily command, 04:10	In Pusula-API the test, answers and result; in Rehnüma the bridge record and e-mail	`pruneStale`, `pusula:prune-links`	Implemented
Pusula-API safety net	Daily scheduled task	Incomplete tests older than 30 days	`Pusula-API/routes/console.php`	Implemented

Path	Trigger	What is deleted	Code	Status
Deletion of the account	User (`DELETE /account`)	All rows attached to the account, the profile, the feedback; the test in Pusula-API and the bridge record in the same request, or in the nightly pruning if the service cannot be reached; order records remain	`AccountSecurity::delete` (ADR-067)	Implemented
The user deleting only their test	User	Test and bridge record	Pusula-API endpoint exists; no Rehnüma endpoint	Missing; the trust page shows the route of a request by e-mail, the button is planned
Deletion of analysis exports	Researcher	JSONL files	`ExportResponses`	No policy

Note. Source: the code in the repository, 22 November 2026.

8.4 Licences

The licence assessments in this section are based on the note on additional dimensions and licences and on the legal approval table of the psychometrics report; both state explicitly that they are not legal advice. The status of the instruments that are used and not used is given in Table 8.4.

Table 8.4. Instrument and data licences, obligations and their counterparts in the implementation.

Instrument or data	Licence provision	Obligation	Counterpart in the implementation	Open task
IPIP-NEO-120 items	Public domain; free for any use, including commercial use	None (attribution is good practice)	Attribution to Johnson (2014) on the entrance screen and at `/analysis/changes`	Personality percentiles are calculated from Johnson's open data (ADR-082); the conditions of use of the data should be assessed separately
O*NET Interest Profiler Short Form	O*NET Tools Developer License; royalty-free, no bar to commercial use	Attribution, copyright notice, indication of changes (§3(a)); in the event of modification, a validation study before release and a "has not approved" notice (§3(b))	English notice in the collapsible "O*NET · IPIP" box on the entrance screen; all 60 items at `/analysis/changes`	Validation study; README notice; validation of the fr/de/ar translations
O*NET Work Importance Locator	Same licence; the instrument was retired on 3 June 2024, "as-is"	Same	All 20 cards and the two adapted cards are in the list	Same
O*NET Database (occupations, Work Styles)	CC BY 4.0	Attribution, link to the licence, indication of changes; trademark statement	At `/analysis/changes`, that occupation matching is done with O*NET 31.0 data; in the `_note` and `source` fields of `occupations.json`	The CC BY / CC BY-ND inconsistency on the licence page

Instrument or data	Licence provision	Obligation	Counterpart in the implementation	Open task
rMEQ / MEQ	Unclear	Written permission from the rights holders	Not used; replaced by a 3-question "rhythm" preference	Reassessment if permission is obtained
Four-letter code, "MBTI", "Myers-Briggs"	The trademarks are registered; the status of the letter codes could not be verified	Trademark search, disclaimer	"Not related to Myers-Briggs" note; original texts (ADR-028 §6)	TÜRKPATENT and EU trademark search
VIA, Grit-S, BFI-2, HEXACO, SWLS	Closed to commercial use or subject to permission	—	Not used (ADR-025, rejected alternative)	—
Enneagram (RHETI), Corpitall content	Copyrighted or unclear	—	Removed (ADR-025 Decision 3; ADR-026 §3)	Documenting that it has been deleted

Note. Source: the note on additional dimensions and licences; the psychometrics report, "Points requiring legal approval" table; `Front-end/src/app/[locale]/(marketing)/analysis/changes/page.tsx` (route names became English with ADR-055; `/test/degisiklikler` in Version 1.1 is `/analysis/changes` today). This is not legal advice.

8.4.1 IPIP: public domain

The IPIP permission statement is clear: because the IPIP has been placed in the public domain, anyone may use IPIP items, scales and inventories for any purpose, commercial or not; permission has been granted in advance (IPIP, n.d.-a). This is one of the main reasons why ADR-025 chose the IPIP-NEO-120: according to ADR-026, the IPIP-NEO-120 is the only facet scale open to commercial use, whereas the BFI-2 and HEXACO are closed. The public domain status applies to the item texts. The Turkish translation is Rehnüma's own; the 924-item Turkish IPIP pool of Somer, Korkmaz and Tatar was used for comparison, but the item texts were not copied from this pool (Chapter 3). The items of Turkish scales such as the 5FKE must not be copied (psychometrics report). The public domain status covers the items; the personality percentiles, by contrast, began with ADR-082 to be calculated from Johnson's (2014) IPIP-NEO-120 open data (619,150 people), and the norm file carries its source with its reference (`Pusula-API/database/data/norms_ipip_neo120.json`, `source`). The conditions of use of these data were not examined separately in the research notes.

8.4.2 The O*NET Tools Developer License and the §3(b) notice

The licence of the O*NET tools (O*NET Resource Center, n.d.-b) grants a worldwide, royalty-free, non-sublicensable, non-exclusive and irrevocable right of use (§2(a)(1)). It requires the user to retain the creator attribution, the copyright notice, the licence reference and the disclaimer of warranties, and to indicate changes (§3(a)). The critical provision is §3(b): if the purpose of the tool is changed, or only part of the material is used, or the content is modified, a validation study must be carried out before the product is released; the study must comply with the *Standards for Educational and Psychological Testing* and 41 C.F.R. Part 60-3. The same provision requires a text informing the end user that USDOL/ETA "has not approved, endorsed, or tested" the modifications to be displayed prominently on the opening screen and in the README (§3(b)(3)). The licence prohibits implying endorsement and grants no trademark rights (§2(a)(6)).

Rehnüma's use triggers §3(b) in more than one way: the items have been translated into Turkish, a cultural adaptation has been made in ten interest items, only the 30-item Mini-IP subset is asked in the first stage, and the wording of two WIL cards has been simplified (note on additional dimensions and licences: in practice the Turkish translation already triggers these two obligations). The notice has been implemented in two places. At the bottom of the entrance screen, inside a collapsible box headed "O*NET · IPIP", an English paragraph says that the product includes information from the O*NET Interest Profiler and Work Importance Locator, that Rehnüma has modified this information (Turkish adaptation, item subset) and that USDOL/ETA has not approved, endorsed, or tested these modifications (`entrance/Entrance.tsx`). The box opens closed; whether it meets the "prominently" condition of §3(b)(3) should be left to legal review. The `/analysis/changes` page repeats the same notice with explicit reference to the O*NET Tools Developer License and gives the full list of changes: the original English text of the 120 IPIP items, the 60 interest items and the 20 value cards, their Turkish equivalent, the stage at which each is asked and whether it has been adapted (`modifications-data.ts`). The page also records the change to the response scale (the move from an accuracy axis to a fit axis), the rewriting of the O6 facet, the translation method and the version stamp (ADR-029 §3; ADR-031 §5). The notice is also held in the data files: the `_note` fields of the `interest_items_onet.json` and `work_values.json` files state the licence and that USDOL/ETA has not approved the modifications.

The five-language decision (ADR-046) widened the scope of the modification. The English items are the original text; the French, German and Arabic items are new translations and have not been validated (ADR-046 §6). This is declared in two places: the version stamp of the changes page in that language says that the translation has not yet been validated and the table shows the item texts of that language (`modifications-items.<dil>.json`); in Pusula-API the `_note` field of the language-specific item files states that the translation has not undergone native-speaker review or psychometric study (`database/data/fr|de|ar/`). With respect to §3(b), these three languages separately give rise to the same validation obligation as the Turkish adaptation (inference).

Two gaps remain in the implementation. The first is the README: §3(b)(3) requires the notice to be present in the README as well. The Pusula-API directory contains only agent instructions (`AGENTS.md`, `CLAUDE.md`), and these do not contain the notice text. Because there is no distributed software package, how the README provision is to be met in a web product requires interpretation; the safest course is to place the notice both in a README at the root of the repository and on the `/analysis/changes` page (inference).

The second, and the weightier, is the validation study. The licence requires a validation study in conformity with the Standards **before the release** of the modified tool and does not define a minimum sample. ADR-029, by contrast, chose to release the product while remaining honest through the language of "first impression" and "technical appendix", and to collect Gate 1–4 data from real users; not releasing until validity evidence arrives is among the rejected alternatives. There is a tension between these two requirements. One solution is for Gate 1 and Gate 2 (internal consistency, item analysis, factor structure and the RIASEC hexagon order) to be completed with a closed pilot before the product is opened to a wide audience, and for this work to be documented as the §3(b) validation study. Whether this solution is sufficient with respect to the licence has been left to review by a lawyer (psychometrics report, legal approval table; ADR-026, "Awaiting legal approval"). For this reason the product texts do not merely avoid the expression "validated"; they also state openly that the validation

study required by the licence has not yet been completed: the changes page writes that the Turkish version "has not yet been verified by an independent study" and "will be updated as the verification gates proceed".

Marketing language is also connected with the licence. Because of the prohibition on implying endorsement, expressions such as "adapted from the O*NET work values framework" are recommended in product text in place of using the O*NET name like a brand (note on additional dimensions and licences). The use of the ® sign and of the instrument names on the changes page is likewise among the points requiring review by a lawyer in the legal approval table of the psychometrics report.

8.4.3 The WIL and occupation data

The O*NET Work Importance Locator was retired on 3 June 2024; no technical support is provided and the materials remain in the archive "as-is" (O*NET Resource Center, n.d.-c). The instrument nevertheless falls within the scope of the Tools licences. There is no licence obstacle to using a retired instrument, but it has two consequences. First, the instrument should not be expected to be updated or to have its errors corrected. Second, the work values layer is the layer with the lowest retest stability (.59–.66); for this reason the report emphasises only the two most prominent values, and the changes page states this openly.

The occupation matching data come with a separate licence. The O*NET Database is licensed under CC BY 4.0; its conditions are attribution of the source, a link to the licence and indication of changes; the sample attribution text includes the statement "O*NET® is a trademark of USDOL/ETA" (O*NET Resource Center, n.d.-d). The psychometrics report notes that there is an inconsistency between CC BY and CC BY-ND on the licence page; this inconsistency directly affects whether processing the occupation data with Rehnüma's own fit formula (a derivative work) is permitted, and it needs to be clarified. The occupation data are held in Pusula-API and are updated once a year with the `pusula:import-onet` command (ADR-027, Impact). The current catalogue was taken from the O*NET 31.0 database and contains 923 occupations; 910 of them are active, and 13 occupations with no counterpart in Türkiye do not take part in matching (`Pusula-API/database/data/occupations.json`, `_note` and `source` fields). The source and the licence are written in the file's own note; it is also stated there that the occupation names and descriptions have been adapted into Turkish.

8.5 Health and finance limits

Character analysis is not a product on its own; its result lives alongside Rehnüma's health, finance and goal modules and can influence their language through the character trace. This proximity creates the risk that the test result turns into clinical or financial guidance. The product has drawn three limits against this risk. The limits are summarised in Table 8.5.

Table 8.5. Health, finance and character trace limits.

Area	What the system does	What the system does not do	Basis
Character analysis	Describes personality, interest, value and rhythm tendencies	Does not diagnose; does not take the place of clinical assessment, career counselling or a hiring decision	Trust page, "What we do not claim"; ITC (2005) 29.1

Area	What the system does	What the system does not do	Basis
Health page	Shows the plan	Contains no recording or entry flow; tracking is in Habits	Project rule; ADR-017 §4
Medication plan	Shows the times	Does not recommend dose, use or interactions	ADR-017 §5
Budget and market	Shows the figures, what remains and the overrun	Does not give an amount, target or investment recommendation according to personality; does not hide any figure	ADR-024; ADR-033 §2; evidence map §4.2
Character trace	When quality is `ok`, softens the language, changes order and emphasis, suggests a time according to Rhythm	Does not hide information, does not change a value, does not build a rule on motif, temperament, archetype, tendency or work value	ADR-033 §1–3; `InvariantsTest`

Note. The character trace is detailed in Chapter 6; here it is summarised only with respect to the ethical limit.

No diagnosis. The trust page states openly that the analysis is not a diagnostic instrument. One consequence of this limit is how the report addresses a person with a low score in the Emotional stability domain. ITC (2005) guideline 30.9 accepts that, in feedback given over the internet, not knowing the emotional state of the test taker and not being able to give immediate support in the event of a negative impact is an ethical problem, and requires that, where appropriate, the feedback include how to reach sources of support. The full report plan of the psychometrics report envisages a separate "Support" section for this purpose: in the band of high anxiety or low emotional stability, a referral to the school guidance service, the family physician and, in an emergency, 112. In the code review this referral was not found in the report; the `report_texts.json` texts describe low Emotional stability in bipolar language but do not show a source of support. This is a gap that needs to be closed, especially for users under 18.

No financial advice. The budget module takes seriously the finding that applications which do not tell the person what to do are abandoned (ADR-024), but it does not answer this "what should I do" question according to personality. The evidence map gives two reasons for this. The first is the smallness of personality effects: relationships of the order of $r \approx .10$–$.20$ between Conscientiousness and exercise and $r = .15$ with medication adherence have been reported (Chapter 6); a band on its own does not say who will give up. The second is that hiding information was helpful in no study: making access to financial information easier reduced fees by 38.4% (Carlin et al., 2023), and showing bad news reduced spending (D'Acunto et al., 2019). Loss framing, for its part, increased anxiety and did not change intention (Dorison et al., 2022). For this reason the overrun figure is not hidden even for low Emotional stability; only tone and order change. The market widget shows price and change; it does not produce a risk appetite or a buy–sell recommendation according to personality. The association of personality data with investment decisions is recorded as `no_change` in the evidence map.

The principle of not hiding information. The most important ethical provision of ADR-033 is that a personality-dependent rule can be a rule only when it reduces or softens; every idea that changes a value or increases something is an A/B candidate that is born closed. The reason is the harm that a mislabelled user would suffer: in Gate 0, run with form v4, the derived labels changed at retest at a rate of 43–55% (ADR-029). Softened language is harmless when applied to the wrong person; a different target given to the wrong person, by contrast, is a wrong plan. The evidence policy also sets an order of precedence for conflicts: the user setting overrides the rule, measured behaviour overrides Rhythm, and an unconditional rule overrides a personality rule. No personality rule can produce the opposite of an unconditional rule; for no band is loss framing, broken-streak emphasis or the hiding

of information opened (Character trace evidence map report, §5.4). This principle is in executable form in `InvariantsTest`: character trace rules cannot be tied to motif, temperament, archetype, tendency or work value.

8.6 The commercial model and ethical tensions

ADR-025 had left the payment step and the question of what would be sold to a separate decision; when Version 1.1 was being written there was no payment step. These decisions were taken later. ADR-041 established the sales model, ADR-062 introduced purchasing on the site (the payment institution is PayTR; card details do not enter Rehnüma's page or server, and purchasing requires sign-in), ADR-073 tied the result to the account, and ADR-083 set the current tiers. Today there are two tiers. The analysis is free and gives the open sections of the report: At a glance, Personality, Interests, Occupations (the first three) and Closing; from each of the nine locked sections other than Five branches, an opening sentence is visible (ADR-081 §7). All sections of the report are opened only with the annual dashboard (membership); the list price of the dashboard is TL 1,500 per year, the launch price is TL 1,200, and the full report is not sold separately (ADR-083). The price and currency are held not in the code but in the admin panel (ADR-063), so the figures here reflect the decision of 22 November 2026. The arrival of the payment step has taken the tensions out of the realm of assumption. The principal ones are gathered in Table 8.6.

Table 8.6. Ethical tensions arising from the commercial model.

Tension	Current decision	Open question
Charging for information derived from the person's own data	The analysis is free; the five open sections are seen with an account, the nine locked sections are opened with the annual dashboard (ADR-081 §7; ADR-083)	Stating the free part and the account condition before the test
Tying the result to the account	The result is not shown without an account; e-mail and password (or sign-in with Google) are requested after the sections (ADR-073)	The server still offering the unregistered summary endpoint; keeping the promise that only transactional messages will be sent to the e-mail (ADR-064)
Requesting an account after effort	The account is requested after 12 minutes of effort; there is no way of seeing the result without opening an account	The sunk-cost feeling forcing the account decision
Display of locked content	The locked section opens: a real opening sentence from the person's own result, followed by blurred filler text that is not real text (ADR-081 §5–6)	The filler being taken for text belonging to the person; its effect on conversion has not been measured
Persuasion according to personality	Personalisation by temperament and motif was rejected; softening only (ADR-033)	The opening criteria for A/B candidates
Adolescent user and payment	Declaration of parental consent; purchasing requires sign-in (ADR-062)	From whom payment will be taken; review of the distance-selling and refund texts in the light of the lira price (ADR-083)
Charging a job-seeking user	The life stage has a "searching" option	The effect on career recommendations of the legislation on charging job seekers

Note. The question in the last row arose from the footer in the competitor products note in which Kariyer.net states that, pursuant to Law No. 4904, no fee may be charged to job seekers; whether it applies to Rehnüma has been left to legal review.

E-mail, account and effort. ADR-026 §4 asked for the e-mail immediately before the summary; according to vendor data this is the position that yields the highest consent rate (Chapter 5.2). ADR-073 removed the e-mail screen and the unregistered summary: when the sections are finished the account step comes directly, and scoring is carried out after the account is opened. The position is the same; it is the moment at which the person has invested 12 minutes of effort and is curious about the result. The bargaining power, however, has increased, because now no result is seen without opening an account. The literature on the labour illusion and the IKEA effect shows that effort increases perceived value; but the IKEA effect arises only when the task is completed (Norton et al., 2012), and when an unsuitable result was shown, foregrounding the effort increased dissatisfaction (Buell & Norton, 2011). The way not to abuse this power is to say honestly what the account is requested for, and to say it before the effort is made as well. The account step says "Your report and your dashboard open in this account" and states that a verification link will be sent; ADR-073, for its part, records in its Impact section, as a separate task, the removal of expressions of the "see your result without signing up" kind from the promotional texts. Until this clean-up is completed, the account condition for the result cannot be regarded as having been stated openly before the test (inference).

Avoiding manipulative design. ADR-031 also defines the trust design in negative terms: there is no marketing cliché, no fabricated social proof, no urgency and no blurred fake content. ADR-081 rebuilt the last provision for locked sections: the locked section opens, the person reads a real opening sentence from their own result, and the rest is blurred. The blurred part is not the person's text; the server sends only the first sentence of each locked section, and the blurred text that is visible is filler built from the promotional sentences of the other sections (ADR-081 §6). The reason is technical and honest: sending the whole of the real text and blurring it in the display could be read with browser tools. The remaining ethical question is whether the filler gives the impression of being text that belongs to the person; unless the interface says that the single real sentence is real and the rest is filler, the tension with ADR-031's "no fake content" principle persists (inference). The drop-off points in the flow have been resolved with text, not with manipulation: the heading of the account step is "Create your account to see your result" and its button is "Create account and see the result", and both rest on the task, not on fear; on the entrance screen of the second stage there is a "Not now, back to my report" route (ADR-031 §4). The "subscription trap" complaints seen in competitor products (2.5/5 on Trustpilot for 16Personalities; competitor products note) suggest that a model with clear conditions should be preferred to a subscription (inference). The current model is in this direction: the dashboard is paid for annually, the payment does not renew automatically, and it is refunded within 30 days (terms of sale page). The same page still speaks of a "full report" sold separately; ADR-083 records the review of the legal sales texts in the light of the two tiers as an open task.

Persuasion according to personality. It has been shown that tailoring messages with personality data increases the persuasive effect on digital audiences (Matz et al., 2017; Hirsh et al., 2012); these findings and their methods have also been criticised (Sharp et al., 2018). Rehnüma's decision is to use this technique in a limited way even for the person's own goals: temperament- and motif-based personalisation ("meaning for the NF, mechanism for the NT") was rejected because it is without evidence and a marketing cliché (ADR-033, rejected alternatives). Personality data should not be used on the payment screen or in the offer text; this is not an ADR provision today (ADR-062 and ADR-083 contain no provision on this subject), and we recommend that it be written into the sales decisions.

The "validated" claim. In a paid product, "validated" or "scientific" marketing claims carry both ethical and legal risk before the validation study is finished (psychometrics report, legal approval table). The avoidance of these expressions on all of Rehnüma's surfaces (ADR-029, ADR-031 §2) is all the more binding now that the payment step has been added (ADR-062).

8.7 Algorithmic fairness

The test sorts people into categories and suggests occupations; when these two operations affect groups systematically differently, a problem of fairness arises. ITC (2005) guideline 29.2 requires the freedom of automated reports from systematic bias to be evaluated. The validation protocol ties this evaluation to Gate 4: a differential item functioning (DIF) analysis for gender and age with at least 200 people per group, and measurement invariance in the 14–17 age group (Chapter 4). The risks, the current measures and the measurement gates are given in Table 8.7.

Table 8.7. Algorithmic fairness risks.

Risk	Evidence	Current measure	Measurement
Gender difference in interest items	O*NET IP-SF: d = .86 in favour of men on Realistic, d = −.59 in favour of women on Social; in Türkiye machine items were perceived as "men's work" (Süerdem & Erkök, 2016)	The Mini-IP's item selection, which minimises the gender difference; gender-neutral language	Item mean by gender for R:02, R:05, R:09 (pilot); DIF (Gate 4)
Reproduction of stereotypes in occupation suggestions	A combined norm reproduces the gender stereotype, a separate norm changes the comparison (validation protocol report)	"Occupational fit is not a prediction of ability or success" note; no percentile is given for interest areas, only a 0–100 score (ADR-082 §5)	The norm decision for interests will be written into an ADR
Comparison of adolescents with adults	Mean C and A levels differ in adolescents (scoring and norms note, S7)	None; personality percentiles come from an international sample not separated by age and gender (ADR-082)	Age-banded norm; 14–17 invariance (Gate 4)
Acquiescence bias in adolescents	Unbalanced keying distorts the structure in adolescents (Soto et al., 2008)	Acquiescence index and `acquiescent` flag; no score correction (ADR-030)	The ADR is updated if the proportion of those aged 14–18 with an absolute index of 1.0 or above exceeds 10% (Gate 1)
Careless responding in adolescents	14–37% flagged among high school students (Chauliac et al., 2023)	No blocking; non-accusatory language	Flag rate < 20% (Gate 1)
Readability	Jargon and low-frequency words for age 14	34 IPIP items, 10 interest items and 2 WIL cards were rewritten (ADR-029 §3)	Cognitive interviewing (before Gate 1)
Accessibility	Screen reader and keyboard walkthrough were not carried out on a real device (ADR-031 Impact)	Accessible accent tone, `aria` labels, reduced motion	Keyboard walkthrough on a real device, 200% zoom

Note. The d values are reported from the psychometric report of the O*NET Interest Profiler Short Form (validation protocol report).

8.7.1 Gender: data in order to measure, a principle in order not to collect

For a gender DIF analysis at least one thing is necessary: the gender of the participant. The code review showed that the test does not ask for gender; neither `StartPersonalityTestRequest` nor Pusula-API's test table contains such a field. This is consistent with the principle of data minimisation and also ensures of itself that the report does not use norms by gender: the published norm on which the personality percentiles rest is likewise not separated by gender (ADR-082, known limitations). However, the same decision makes the gender DIF analysis envisaged by Gate 4, and the item "gender DIF monitoring (R domain)" in ADR-029's list of open tasks, impossible to carry out. Given that large gender differences in interest items are known (Table 8.7), it cannot be known, without this analysis being done, whether the occupation suggestions reproduce gender stereotypes.

There is a real tension here between data protection and the fairness audit. Three options can be seen: (a) not asking for gender at all and giving up the DIF analysis; (b) asking for gender optionally, for aggregate analysis only and with an explicit statement that it will never be used in the report; (c) carrying out the analysis only in a pilot group that gives separate consent. Options (b) and (c) can be set up in a way that is compatible with purpose limitation, but they require a separate privacy notice and consent. The decision should be taken through an ADR; this chapter recommends option (c) as the least costly route, because the Gate 1 pilot is in any case to be conducted with a separate consent (inference).

8.7.2 Age

Age, unlike gender, is collected: the year of birth exists in both services and the export writes the age. Age analysis is therefore technically possible. The problem lies in reporting. Personality percentiles now come not from the reference group of those who have taken the test in Rehnüma but from the published international sample of the IPIP-NEO-120 (Johnson, 2014; 619,150 people). This sample includes every age and country together; no separate norm by age and gender has been used, and the norm is based on volunteers who completed the English form online, so that carrying it over to the Turkish adaptation is an assumption (ADR-082, known limitations; Chapter 5.4). Considering that the Conscientiousness and Agreeableness means of adolescents differ from those of adults, the percentile of a 15-year-old user may come out systematically low in a group in which adults predominate. When this percentile enters a rule in the character trace, for example when a smaller step is suggested in the low Conscientiousness band, an age-related difference in maturation is treated as if it were a personality trait. The character trace's rules, which only soften, limit this harm, but the label itself remains in the report. The language of "not yet", "at present", "in this period" and the framing of "areas worth exploring" that the scoring note recommends for adolescents may reduce the effect of this labelling; this language has not yet been applied systematically in the report. When Rehnüma's own Turkish norm has been collected (`ReferenceGroup` remains in the code but is not connected to the report; ADR-082), modelling the 14–17 and 18+ groups separately is the explicit recommendation of the validation protocol (EFPA, 2025; no extrapolation is made beyond the observed age range).

8.7.3 Readability and language

The lower end of the target audience is the age of 14, and the items were reviewed for this age. The translation and item quality review identified a lexical or morphological load for age 14 in approximately 12 of the IPIP items (e.g. "ayartma" [temptation], "kuramsal" [theoretical]) and showed that in

the interest items the main problem was the office and finance jargon in the Enterprising and Conventional areas ("tahvil" [bond], "müvekkil" [client of a lawyer], "el terminali" [handheld terminal]). On the basis of these findings, 34 IPIP items, 10 interest items and 2 WIL cards were rewritten before the pilot (ADR-029 §3). No formal Turkish readability index was calculated for readability; the review rests on expert judgement. Cognitive interviewing and a second independent translation are planned for before Gate 1 (`/analysis/changes`). This review is for the Turkish items only; the French, German and Arabic item translations have not undergone readability or native-speaker review (ADR-046 §6).

One aspect of language that bears on fairness is the response scale. A change of labels can change the meaning of the scores: with different midpoint labels on the same items, only configural invariance was achieved and metric invariance could not be achieved (Doğan et al., 2024; validation protocol report). For this reason the label change was made before the pilot, and it was decided that the labels would not change after the pilot. The rewriting of the items in the O6 (Liberalism) facet that are tied to US politics is likewise a fairness decision, both in terms of cultural appropriateness and in terms of not asking a 14-year-old user about political preference.

Finally, accessibility is a matter of fairness: if a measurement instrument is harder to complete for someone who uses a screen reader or has motion sensitivity, this affects the quality of that person's answers and therefore their report. ADR-031 §7 set the accessibility thresholds but left the keyboard walkthrough on a real device, the screen reader, 200% zoom and iOS checks as open tasks. Showing the five labels on the Likert screen beneath the circles on a wide screen and at the two ends on a narrow screen, and having the attention-check items ask for the option both by its name and by its number, aim to equalise the instruction for users in different screen and reading conditions (ADR-026 addendum).

Chapter references

- Buell, R. W., & Norton, M. I. (2011). The labor illusion: How operational transparency increases perceived value. *Management Science, 57*(9), 1564–1579. https://doi.org/10.1287/mnsc.1110.1376
- Carlin, B., Olafsson, A., & Pagel, M. (2023). Mobile apps and financial decision making. *Review of Finance, 27*(3), 977–996. https://doi.org/10.1093/rof/rfac040
- Chauliac, M., Willems, J., Gijbels, D., & Donche, V. (2023). The prevalence of careless response behaviour and its consequences on data quality in self-report questionnaires on student learning. *Frontiers in Education, 8*, 1197324. https://doi.org/10.3389/feduc.2023.1197324
- D'Acunto, F., et al. (2019). *Status app peer-spending feedback* [Chicago Booth Review summary]. https://chicagobooth.edu/review/how-crowdsourcing-app-can-stop-people-overspending (reference to be verified)
- Doğan, Gündeğer, & Yurtçu. (2024). [Measurement invariance study with four different midpoint labels on the same 14 items]. https://acikerisim.aksaray.edu.tr/items/66f7f62d-cc6d-4fa5-95e9-fd7cbfbc20a6/full (reference to be verified)
- Dorison, C. A., et al. (2022). In COVID-19 health messaging, loss framing increases anxiety with little-to-no concomitant benefits: Experimental evidence from 84 countries. *Affective Science, 3*(3), 577–602. https://doi.org/10.1007/s42761-022-00128-3

- EFPA. (2025). *EFPA review model for the description and evaluation of psychological and educational tests* (Version 2025.5). European Federation of Psychologists' Associations. https://www.infocop.es/wp-content/uploads/2025/09/efpa_test_review_model_v2025_5.pdf
- Hirsh, J. B., Kang, S. K., & Bodenhausen, G. V. (2012). Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits. *Psychological Science, 23*(6), 578–581. https://doi.org/10.1177/0956797611436349
- International Test Commission. (2005). *International guidelines on computer-based and internet-delivered testing*. https://www.intestcom.org/files/guideline_computer_based_testing.pdf
- IPIP. (n.d.-a). *Permission*. International Personality Item Pool. https://ipip.ori.org/newPermission.htm
- Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003
- Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences, 114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114
- Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology, 22*(3), 453–460. https://doi.org/10.1016/j.jcps.2011.08.002
- O*NET Resource Center. (n.d.-b). *O*NET Tools Developer License*. https://www.onetcenter.org/license_toolsdev.html
- O*NET Resource Center. (n.d.-c). *O*NET Work Importance Profiler and Work Importance Locator*. https://www.onetcenter.org/WIP.html
- O*NET Resource Center. (n.d.-d). *O*NET Database license*. https://www.onetcenter.org/license_db.html
- Sharp, B., Danenberg, N., & Bellman, S. (2018). Psychological targeting [Letter to the editor]. *Proceedings of the National Academy of Sciences, 115*(34), E7890. https://doi.org/10.1073/pnas.1810436115
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of Big Five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718–737. https://doi.org/10.1037/0022-3514.94.4.718
- Süerdem, A., & Erkök, B. (2016). Assessing the reliability and validity of a shorter version of RIASEC in Turkish. *SHS Web of Conferences, 26*, 01063. https://doi.org/10.1051/shsconf/20162601063

9. Discussion

9.1 Answers to the research questions

RQ1 — With which open-source instruments can an ambitious yet honest test be built? Three publicly accessible instruments (the IPIP-NEO-120, the O*NET Interest Profiler short form and the O*NET Work Importance Locator) and a short rhythm self-report offer a broader measurement base than most commercial tests on the market: 30 facets, six interest areas, six work values and chronotype (Chapter 3). The original versions of these instruments have been tested in large samples (Johnson, 2014; O*NET Resource Center). What is "ambitious" is the breadth of measurement; what is "honest" is that each layer is presented at its own level of evidence: continuous scores are primary, derived labels come with an uncertainty marker, and while the evidence for the Turkish version has not yet been collected the report says so.

RQ2 — How are validity and reliability demonstrated in stages? The five-gate protocol (Chapter 4) ties each claim to a condition, a threshold and an output. Gate 0 tested the only thing that can be tested without user data: the correctness of the scoring pipeline and the sensitivity of categorical labels to measurement error. The former passed as expected; the latter revealed two flaws that changed the design (the balanced letter, neck-and-neck interests). The value of the protocol lies not in its findings but in its having been written before the findings arrived: the thresholds are fixed, the language is tied to the gate, and the "validated" claim is closed.

RQ3 — How is the result reported without falling into the Barnum effect? By tying every sentence to a score, writing both poles as strengths, pruning on the basis of feedback and making uncertainty visible (Chapter 5). The competitor review (Chapter 2) showed that patterns such as fake text in locked content, urgency and social proof are widespread; in their place Rehnüma puts a table of contents and a technical appendix; the locked section opens, shows a real opening sentence from the person's own result, and the part of the remainder that appears blurred is not real text but filler built from introductory sentences (ADR-081; the tension is discussed in Chapter 9.4). This is a requirement of measurement ethics as much as it is a product choice.

RQ4 — With what evidence can personality findings be linked to the decisions of a life dashboard? The six-strand evidence map (Chapter 6) gave a narrower answer than expected: not a single RCT could be found showing, with a behavioural outcome, that a message tailored to personality changes behaviour more. The usable evidence is concentrated in two domains (Emotional stability, Conscientiousness) and in rhythm; the strongest product rules, by contrast, are independent of personality. This result redefined the promise "let every widget speak according to personality" as "let every widget soften where the evidence permits". The shrinking of the promise is not a loss but an indication that the evidence policy is working.

9.2 Relation of the findings to the literature

The label instability observed in the simulation is not a new finding; it has long been known that dichotomising a continuous variable produces information loss and instability (MacCallum et al., 2002), and MBTI type-change rates (Pittenger, 2005) are the reflection of the same mechanism in real data. Our contribution is to measure this in our own scoring pipeline, before the user sees it, and to write it back into the design. The main verdict of the evidence map is also consistent with the literature: personality effects are small at the main-effect level (Jokela et al., 2013a; Molloy et al., 2014; Wilson & Dishman, 2015) and tailored-persuasion studies mostly remain at the level of perception (Alslaity et al., 2023). By contrast, personality-independent findings such as framing (Koo & Fishbach, 2012), loss language (Dorison et al., 2022) and streak breaking (Silverman & Barasch, 2023) could be translated directly into product rules.

9.3 Limitations

1. **There is no psychometric evidence for the Turkish version.** The translation was made with a single forward translation and an item review; a second independent translation and cognitive interviews are pending before the pilot. The source language of the product content is Turkish; the English items are the original texts of the instruments, whereas the French, German and Arabic item translations are new and have not undergone native-speaker review (ADR-046). The Gate 0 findings show only the correctness of the code.
2. **The simulation was generated from our own model.** The simulation was run with form v4; the two ranking steps added by form v5 (ADR-044) were not tested in the simulation. The response model of the synthetic respondents and the scoring model share the same assumptions; the recovery values are the upper bound of what will be reached in real data. In addition, the reported label stabilities (four letters 57%) belong to the simulation run with the 6-point balanced threshold; no rerun was made after the threshold was raised to 8, and the raw outputs were not kept in the repository (Chapter 4). It is planned to rerun it before Gate 1 and to place the outputs under `docs/research/sim/`.
3. **There is no Turkish norm.** Personality percentiles are based not on those who take the test in Rehnüma but on the published international sample of the IPIP-NEO-120 (Johnson, 2014; 619,150 respondents; ADR-082). This norm comes from volunteers who completed the English form online, and carrying it over to the Turkish adaptation is an assumption; the sample contains every age and country together, and no separate norms by age and sex were used. Because its three rewritten items are not present in the original data, no percentile is given for facet O6; because there is no published norm table for the interest areas, only the 0–100 score is shown.
4. **The basis of the derived layers is limited.** The eight archetypes are inspired by the clusters of Gerlach et al. (2018), but their centres are the authors' own grouping; the clustering will be redone with Gate 2 data.
5. **The evidence map does not rest on a sample from Turkey.** The only direct Turkish study is at the level of perception (Alkış & Taşkaya Temizel, 2015); cultural generalisability is an assumption.
6. **Single-user development.** The product is developed through dogfooding; without a real user population, A/B candidates can be tested only by comparison across periods, which is a weak design.

7. **Authorship and independence.** Design, translation, analysis and reporting were carried out by the same team; there is no independent evaluation.

9.4 Tensions between product and science

Three tensions are worth documenting. The first is between the wish for a visible and memorable label (ISTP, Meticulous Architect) and the fragility of such labels; the solution has been not to hide the label but to carry the uncertainty (ADR-028, ADR-029). Second, the lock between the free summary and the full report has to arouse curiosity without hiding information; the table of contents and the sample report strike this balance. In the current arrangement the locked section gives a real opening sentence from the person's own result, the remainder is not sent to the client and the lines that appear blurred are filler (ADR-081); this is a choice that comes close to the blurred filler text criticised in competitors in Chapter 2, and whether the filler gives the impression of real content should be monitored through user feedback (inference). All sections of the report are unlocked not as a separate product but with the annual dashboard membership (ADR-083). The third is streak design: the product's "no make-up day" decision (ADR-004) conflicts with the finding that emphasising a broken streak increases abandonment (ADR-033). The decision was softened in this direction by an addendum: there is still no tolerance day, but retroactive marking counts towards the streak (ADR-004 addendum, 25 October 2026).

Chapter references

Alkış, N., & Taşkaya Temizel, T. (2015). The impact of individual differences on influence strategies. *Personality and Individual Differences, 87*, 147–152. https://open.metu.edu.tr/handle/11511/30500

Alslaity, A., Chan, G., & Orji, R. (2023). A panoramic view of personalization based on individual differences in persuasive and behavior change interventions. *Frontiers in Artificial Intelligence, 6*, 1125191. https://doi.org/10.3389/frai.2023.1125191

Dorison, C. A., et al. (2022). In COVID-19 health messaging, loss framing increases anxiety with little-to-no concomitant benefits: Experimental evidence from 84 countries. *Affective Science, 3*(3), 577–602. https://doi.org/10.1007/s42761-022-00128-3

Gerlach, M., Farb, B., Revelle, W., & Amaral, L. A. N. (2018). A robust data-driven approach identifies four personality types across four large data sets. *Nature Human Behaviour, 2*, 735–742. https://doi.org/10.1038/s41562-018-0419-z

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Jokela, M., et al. (2013a). Association of personality with the development and persistence of obesity: A meta-analysis based on individual-participant data. *Obesity Reviews, 14*(4), 315–323. https://doi.org/10.1111/obr.12007

Koo, M., & Fishbach, A. (2012). The small-area hypothesis: Effects of progress monitoring on goal adherence. *Journal of Consumer Research, 39*(3), 493–509. https://academic.oup.com/jcr/article/39/3/493/1795380

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19–40. https://psychology.psy.sunysb.edu/attachment/measures/content/maccallum_on_dichotomizing.pdf

Molloy, G. J., O'Carroll, R. E., & Ferguson, E. (2014). Conscientiousness and medication adherence: A meta-analysis. *Annals of Behavioral Medicine, 47*(1), 92–101. https://doi.org/10.1007/s12160-013-9524-4

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research, 57*(3), 210–221. https://doi.org/10.1037/1065-9293.57.3.210

Silverman, J., & Barasch, A. (2023). On or off track: How (broken) streaks affect consumer decisions. *Journal of Consumer Research, 49*(6), 1095–1117. https://doi.org/10.1093/jcr/ucac029

Wilson, K. E., & Dishman, R. K. (2015). Personality and physical activity: A systematic review and meta-analysis. *Personality and Individual Differences, 72*, 230–242. https://doi.org/10.1016/j.paid.2014.08.023

10. Conclusion and Future Work

10.1 Conclusion

This study has shown how a personality and vocational orientation test built with open-source psychometric instruments can be offered within a consumer product while preserving its scientific honesty. Three principles bind all decisions: every claim stands behind an evidence gate; the continuous score is primary, the label is derivative and is presented with its uncertainty; personalisation cannot hide information, it only changes the language and the default as far as the evidence permits. These principles have also been built into the code: the validation gates protect the report language, the evidence policy protects the rule catalogue, and the invariant tests protect the decision records.

The concrete outputs in hand in the pre-pilot state are the following: a two-stage form with response-quality checks (120 personality items, 60 interest items, 20 value items and 3 rhythm questions; with form v5, self-placement and an eight-round statement ranking, ADR-044); a single-ranked and diverse matching against 910 occupations; a scoring pipeline tested with 300 synthetic respondents (form v4); a fourteen-section report delivered in five languages, with personality percentiles drawn from the published norm (Johnson, 2014; ADR-082) and with uncertainty markers; a public method page and change log; an evidence-graded rule catalogue of 341 records for seventeen surfaces and an engine that applies it; and a character matrix of 76 cells and 209 sources for five life branches (ADR-037).

10.2 Future work

The order cannot be changed, because each step depends on the output of the previous one.

1. **Second independent translation and cognitive interviews.** Before the pilot; 5–10 participants, at least three of them aged 14–17. The interview protocol, participant texts and reconciliation table are ready under `docs/research/pilot/`; the interviews have not yet been conducted.
2. **Gate 1 pilot.** About 100 real users; internal consistency, item analysis, calibration of the carelessness thresholds on Turkish and mobile data, distribution of the acquiescence index.
3. **Gates 2 and 3.** Factor structure and the RIASEC circular order in at least 300 people; retest at an interval of 4–8 weeks; measurement of motif and archetype stability in real data; re-clustering of the archetype centres.
4. **Gate 4.** Norms banded by age and sex, DIF monitoring (the Realistic area in particular), Barnum pruning; the expression "Turkey norm" only after this gate. Until then personality percentiles are given relative to the published international sample (ADR-082).
5. **Applying the character trace to the surfaces.** First the unconditional rules, then the Emotional stability softenings, then the times derived from rhythm; no A/B candidate is opened until the measurement events (`widget_view`, `reminder_*`, `default_changed`, `copy_variant_shown`, `rule_overridden`) are being recorded.
6. **Publication.** If the Turkish adaptation and the Gate 1–3 findings are to be submitted to a psychometrics journal or a preprint server, ethics committee approval and the academic form of the in-

formed consent text will be required; Version 1.0 of this monograph can be used as the pre-registration document of that study.

The next version of the monograph will be published with the Gate 1 findings; previous versions will be preserved unchanged.

Chapter references

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

References

The references at the end of the chapters have been merged and sorted alphabetically (APA 7). Entries marked "(reference to be verified)" are works whose full reference is not available in the research notes and which must be confirmed against the primary source before publication. Repository documents (decision records, research notes, source code) are cited by path in each chapter and are not repeated here.

Repository documents. ADR-003, ADR-004, ADR-010, ADR-013, ADR-016, ADR-017, ADR-019, ADR-024, ADR-025, ADR-026, ADR-027, ADR-028, ADR-029, ADR-030, ADR-031, ADR-032, ADR-033, ADR-034, ADR-035, ADR-037, ADR-038, ADR-041, ADR-044, ADR-045, ADR-046, ADR-055, ADR-058, ADR-061, ADR-062, ADR-063, ADR-064, ADR-065, ADR-067, ADR-073, ADR-074, ADR-080, ADR-081, ADR-082, ADR-083 (`docs/decisions/`); `docs/research/reports/Karakter izi kanıt haritası.md`; `docs/research/research_notes/Karakter izi kanıt haritası/` (six notes); `Front-end/src/widgets/KARAKTER-IZI.md` and the seventeen `KARAKTER.md` files that have a rule catalogue; `Back-end/database/data/personality_rules/`; `Back-end/app/Services/PersonalityRules/`; `Front-end/src/lib/usePersonalityRules.ts`; `Front-end/src/components/test/report/ornek-rapor.json`.

16Personalities. (n.d.-a). *Our theory*. https://www.16personalities.com/articles/our-theory

16Personalities. (n.d.-b). *Premium*. https://www.16personalities.com/premium

16Personalities. (n.d.-c). *Reliability and validity*. https://www.16personalities.com/articles/reliability-and-validity

Adan, A., & Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: A reduced scale. *Personality and Individual Differences, 12*(3), 241–253. https://doi.org/10.1016/0191-8869(91)90110-W

Ajzen, I., Czasch, C., & Flood, M. G. (2009). From intentions to behavior: Implementation intention, commitment, and conscientiousness. *Journal of Applied Social Psychology, 39*(6), 1356–1372. https://doi.org/10.1111/j.1559-1816.2009.00485.x

Alarcon, G., Eschleman, K. R., & Bowling, N. A. (2009). Relationships between personality variables and burnout: A meta-analysis. *Work & Stress, 23*(3), 244–263. https://doi.org/10.1080/02678370903282600

Albertson, E. R., Neff, K. D., & Dill-Shackleford, K. E. (2015). Self-compassion and body dissatisfaction in women: A randomized controlled trial. *Mindfulness, 6*(3), 444–454. https://mijn.bsl.nl/self-compassion-and-body-dissatisfaction-in-women-a-randomized-c/653090

Alderotti, G., Rapallini, C., & Traverso, S. (2023). The Big Five personality traits and earnings: A meta-analysis. *Journal of Economic Psychology, 94*, 102570. https://doi.org/10.1016/j.joep.2022.102570

Algoe, S. B., Gable, S. L., & Maisel, N. C. (2010). It's the little things: Everyday gratitude as a booster shot for romantic relationships. *Personal Relationships, 17*(2), 217–233. https://doi.org/10.1111/j.1475-6811.2010.01273.x

Alkış, N., & Taşkaya Temizel, T. (2015). The impact of individual differences on influence strategies. *Personality and Individual Differences, 87*, 147–152. https://open.metu.edu.tr/handle/11511/30500

Allen, M. S., & Robson, D. A. (2020). Personality and body dissatisfaction: An updated systematic review with meta-analysis. *Body Image, 33*, 77–89. https://doi.org/10.1016/j.bodyim.2020.02.001

Alslaity, A., Chan, G., & Orji, R. (2023). A panoramic view of personalization based on individual differences in persuasive and behavior change interventions. *Frontiers in Artificial Intelligence, 6*, 1125191. https://doi.org/10.3389/frai.2023.1125191

Alwin, D. F., & Krosnick, J. A. (1985). The measurement of values in surveys: A comparison of ratings and rankings. *Public Opinion Quarterly, 49*(4), 535–552. https://doi.org/10.1086/268949

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/open-access-files.html

American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. https://www.ucc.ie/en/media/academic/appliedpsychology/ethicsforms/apa_code2002.pdf

American Thoracic Society. (n.d.). *Munich ChronoType Questionnaire (MCTQ)*. https://site.thoracic.org/assemblies/srn/sleep-related-questionnaires/mctq

Ames, D. R., & Flynn, F. J. (2007). What breaks a leader: The curvilinear relation between assertiveness and leadership. *Journal of Personality and Social Psychology, 92*(2), 307–324. https://doi.org/10.1037/0022-3514.92.2.307

Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology, 74*(6), 1531–1544. https://doi.org/10.1037/0022-3514.74.6.1531

Atak, H. (2013). On Maddeli Kişilik Ölçeği'nin Türk kültürüne uyarlanması. *Nöropsikiyatri Arşivi, 50*(4), 312–319. https://doi.org/10.4274/npa.y6128

Atlı, A., & Keldal, G. (2017). [Vocational Personality Types Inventory: validity and reliability in high-school students]. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 18*(1), 73–93. https://dergipark.org.tr/tr/pub/ogusbd/article/330743 (reference to be verified)

Barańczuk, U. (2019). The Five Factor Model of personality and social support: A meta-analysis. *Journal of Research in Personality, 81*, 38–46. https://doi.org/10.1016/j.jrp.2019.05.002

Baranski, E., et al. (2024). Personality and conceptions of religiosity across the world's religions. *Journal of Research in Personality, 110*, 104496. https://doi.org/10.1016/j.jrp.2024.104496

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium. *International Journal of Selection & Assessment, 9*(1–2), 9–30. https://doi.org/10.1111/1468-2389.00160

Bartram, D. (2005). The Great Eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*(6), 1185–1203. https://pubmed.ncbi.nlm.nih.gov/16316273

Bazley, W. J., Cronqvist, H., & Mormann, M. (2021). Visual finance: The pervasive effects of red on investor behavior. *Management Science, 67*(9), 5616–5641. https://papers.ssrn.com/ab_id=2992812

Benn, Y., Webb, T. L., Chang, B. P. I., & Harkin, B. (2016). What is the psychological impact of self-weighing? A meta-analysis. *Health Psychology Review, 10*(2), 187–203. https://doi.org/10.1080/17437199.2016.1138871

Berry, D. S., & Hansen, J. S. (2000). Personality, nonverbal behavior, and interaction quality in female dyads. *Personality and Social Psychology Bulletin, 26*(3), 278–292. https://doi.org/10.1177/0146167200265002

Bess, T. L., & Harvey, R. J. (2002). Bimodal score distributions and the MBTI: Fact or artifact? *Journal of Personality Assessment, 78*(1), 176–186. https://doi.org/10.1207/S15327752JPA7801_11

Bidargaddi, N., et al. (2018). To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. *JMIR mHealth and uHealth, 6*(11), e10123. https://doi.org/10.2196/10123

Billieux, J., Rochat, L., Rebetez, M. M. L., & Van der Linden, M. (2008). Are all facets of impulsivity related to self-reported compulsive buying behavior? *Personality and Individual Differences, 44*(6), 1432–1442. https://doi.org/10.1016/j.paid.2007.12.011

Bloch, L., Haase, C. M., & Levenson, R. W. (2014). Emotion regulation predicts marital satisfaction: More than a wives' tale. *Emotion, 14*(1), 130–144. https://doi.org/10.1037/a0034272

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin, 130*(6), 887–919. https://doi.org/10.1037/0033-2909.130.6.887

Breines, J. G., & Chen, S. (2012). Self-compassion increases self-improvement motivation. *Personality and Social Psychology Bulletin, 38*(9), 1133–1143. https://pubmed.ncbi.nlm.nih.gov/22645164/

Briley, D. A., & Tucker-Drob, E. M. (2012). Broad bandwidth or high fidelity? Evidence from the structure of genetic and environmental effects on the facets of the Five Factor Model. *Behavior Genetics, 42*(5), 743–763. https://doi.org/10.1007/s10519-012-9548-8

Brown, S. D., & Ryan Krane, N. E. (2000). Four (or five) sessions and a cloud of dust: Old assumptions and new observations about career counseling. In S. D. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (3rd ed.). Wiley. https://ncda.org/aws/NCDA/pt/sd/news_article/5310/_PARENT/CC_layout_details/false (reference to be verified)

Brown, S., & Taylor, K. (2014). Household finances and the "Big Five" personality traits. *Journal of Economic Psychology, 45*, 197–212. https://eprints.whiterose.ac.uk/82777/3/WRRO_82777.pdf

Brühlmann, F., Memeti, Z., Aeschbach, L. F., Perrig, S. A. C., & Opwis, K. (2024). The effectiveness of warning statements in reducing careless responding in crowdsourced online surveys. *Behavior Research Methods, 56*(6), 5862–5875. https://doi.org/10.3758/s13428-023-02321-z

Brummelman, E., et al. (2014). "That's not just beautiful—That's incredibly beautiful!" The adverse impact of inflated praise on children with low self-esteem. *Psychological Science, 25*(3), 728–735. https://doi.org/10.1177/0956797613514251

Buecker, S., Maes, M., Denissen, J. J. A., & Luhmann, M. (2020). Loneliness and the Big Five personality traits: A meta-analysis. *European Journal of Personality, 34*(1), 8–28. https://doi.org/10.1002/per.2229

Buell, R. W., & Norton, M. I. (2011). The labor illusion: How operational transparency increases perceived value. *Management Science, 57*(9), 1564–1579. https://doi.org/10.1287/mnsc.1110.1376

Buss, D. M. (1991). Conflict in married couples: Personality predictors of anger and upset. *Journal of Personality, 59*(4), 663–688. https://doi.org/10.1111/j.1467-6494.1991.tb00926.x

Carlin, B., Olafsson, A., & Pagel, M. (2023). Mobile apps and financial decision making. *Review of Finance, 27*(3), 977–996. https://doi.org/10.1093/rof/rfac040

Carlo, G., Okun, M. A., Knight, G. P., & de Guzman, M. R. T. (2005). The interplay of traits and motives on volunteering. *Personality and Individual Differences, 38*(6), 1293–1305. https://digitalcommons.unl.edu/psychfacpub/9

Cemalcılar, Z., Baruh, L., Kezer, M., Soto, C. J., Sümer, N., & John, O. P. (2021). Testing the BFI-2 in a non-WEIRD community sample. *Personality and Individual Differences, 182*, 111087. https://doi.org/10.1016/j.paid.2021.111087

CharacterIX. (n.d.). *Mesleki tercih*. https://characterix.global/CareerChoice/TR

Chauliac, M., Willems, J., Gijbels, D., & Donche, V. (2023). The prevalence of careless response behaviour and its consequences on data quality in self-report questionnaires on student learning. *Frontiers in Education, 8*, 1197324. https://doi.org/10.3389/feduc.2023.1197324

Chu, C., Russell, M. T., Hoff, K. A., Phan, W. M. J., & Rounds, J. (2022). What do interest inventories measure? The convergence and content validity of four RIASEC inventories. *Journal of Career Assessment, 30*(4), 776–801. https://doi.org/10.1177/10690727221081554

Churchill, S., & Jessop, D. C. (2011). Too impulsive for implementation intentions? Evidence that impulsivity moderates the effectiveness of an implementation intention intervention. *Psychology & Health, 26*(5), 517–530. https://doi.org/10.1080/08870441003611536

Cianci, A. M., Klein, H. J., & Seijts, G. H. (2010). The effect of negative feedback on tension and subsequent performance: The main and interactive effects of goal content and conscientiousness. *Journal of Applied Psychology, 95*(4), 618–630. https://doi.org/10.1037/a0019130

Ciocarlan, A., Masthoff, J., & Oren, N. (2019). Actual persuasiveness: Impact of personality, age and gender on message type susceptibility. *PERSUASIVE 2019, Lecture Notes in Computer Science, 11433*, 283–294. https://link.springer.com/10.1007/978-3-030-17287-9_23

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation. *Journal of Applied Psychology, 85*(5), 678–707. https://doi.org/10.1037/0021-9010.85.5.678

Conrad, F. G., Tourangeau, R., Couper, M. P., & Zhang, C. (2017). Reducing speeding in web surveys by providing immediate feedback. *Survey Research Methods, 11*(1), 45–61. https://ojs.ub.uni-konstanz.de/srm/article/view/6304 (reference to be verified)

Covey, J. (2014). The role of dispositional factors in moderating message framing effects. *Health Psychology, 33*(1), 52–65. https://durham-repository.worktribe.com/output/1477215

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology, 102*(4), 874–888. https://neoacademic.com/2012/03/21/dont-use-abbreviated-personality-measures/

Curran, P. G., & Hauser, K. A. (2019). I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items. *Journal of Research in Personality, 82*, 103849. https://www.gwern.net/doc/psychology/personality/2019-curran.pdf

D'Acunto, F., et al. (2019). *Status app peer-spending feedback* [Chicago Booth Review summary]. https://chicagobooth.edu/review/how-crowdsourcing-app-can-stop-people-overspending (reference to be verified)

Dai, H., Milkman, K. L., & Riis, J. (2014). The fresh start effect. *Management Science, 60*(10), 2563–2582. https://faculty.wharton.upenn.edu/wp-content/uploads/2014/06/Dai_Fresh_Start_2014_Mgmt_Sci.pdf

Dalton, A. N., & Spiller, S. A. (2012). Too much of a good thing: The benefits of implementation intentions depend on the number of goals. *Journal of Consumer Research, 39*(3), 600–614. https://ideas.repec.org/a/oup/jconrs/doi10.1086-664500.html

Dennis, M., Masthoff, J., & Tintarev, N. (2015). *Adapting reminders to conscientiousness* [PATH 2015 workshop paper]. https://eprints.bournemouth.ac.uk/23268/1/PATH2015-paper6.pdf (reference to be verified)

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Dickson, D. H., & Kelly, I. W. (1985). The "Barnum effect" in personality assessment: A review of the literature. *Psychological Reports, 57*(2), 367–382. https://doi.org/10.2466/pr0.1985.57.2.367

Diener, E. (n.d.). *Satisfaction with Life Scale*. https://eddiener.com/scales/7

Doğan, Gündeğer, & Yurtçu. (2024). [Measurement invariance study with four different midpoint labels on the same 14 items]. https://acikerisim.aksaray.edu.tr/items/66f7f62d-cc6d-4fa5-95e9-fd7cbfbc20a6/full (reference to be verified)

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*(2), 192–203. https://pubmed.ncbi.nlm.nih.gov/16768595/

Donnelly, G., Iyer, R., & Howell, R. T. (2012). The Big Five personality traits, material values, and financial well-being of self-described money managers. *Journal of Economic Psychology, 33*(6), 1129–1142. https://ideas.repec.org/a/eee/joepsy/v33y2012i6p1129-1142.html

Dorison, C. A., et al. (2022). In COVID-19 health messaging, loss framing increases anxiety with little-to-no concomitant benefits: Experimental evidence from 84 countries. *Affective Science, 3*(3), 577–602. https://doi.org/10.1007/s42761-022-00128-3

Driver, J. L., & Gottman, J. M. (2004). Daily marital interactions and positive affect during marital conflict among newlywed couples. *Family Process, 43*(3), 301–314. https://doi.org/10.1111/j.1545-5300.2004.00024.x

Druiven, S. J. M., et al. (2020). Stability of chronotype over a 7-year follow-up period and its association with severity of depressive and anxiety symptoms. *Depression and Anxiety, 37*(5), 466–474. https://doi.org/10.1002/da.22995

Duckworth, A. (n.d.). *Research*. https://angeladuckworth.com/research/

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance. *Journal of Applied Psychology, 91*(1), 40–57. https://doi.org/10.1037/0021-9010.91.1.40

Duggan, K. A., Friedman, H. S., McDevitt, E. A., & Mednick, S. C. (2014). Personality and healthy sleep: The importance of conscientiousness and neuroticism. *PLoS ONE, 9*(3), e90628. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3961248/

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Dyrenforth, P. S., Kashy, D. A., Donnellan, M. B., & Lucas, R. E. (2010). Predicting relationship and life satisfaction from personality in nationally representative samples from three countries. *Journal of Personality and Social Psychology, 99*(4), 690–702. https://pubmed.ncbi.nlm.nih.gov/20718544/

EFPA. (2013). *EFPA review model for the description and evaluation of psychological and educational tests* (Version 4.2.6). https://cms.bps.org.uk/sites/default/files/2022-07/110c_-_efpa_boa_test_review_model_version_4_2_6_approved.pdf

EFPA. (2025). *EFPA review model for the description and evaluation of psychological and educational tests* (Version 2025.5). European Federation of Psychologists' Associations. https://www.infocop.es/wp-content/uploads/2025/09/efpa_test_review_model_v2025_5.pdf

Eggerth, D. E., & Andrew, M. E. (2006). Modifying the C index for use with Holland codes of unequal length. *Journal of Career Assessment, 14*(2), 267–275. https://doi.org/10.1177/1069072705283976

Einarsdóttir, S., & Rounds, J. (2020). A quantitative review of gender differences in vocational interests in Iceland: Pervasive and persistent. *Nordic Journal of Transitions, Careers and Guidance, 1*(1), 10–26. https://doi.org/10.16993/njtcg.29

Elfhag, K., & Morey, L. C. (2008). Personality traits and eating behavior in the obese. *Eating Behaviors, 9*, 285–293. https://europepmc.org/search?query=%22Personality%20traits%20and%20eating%20behavior%20in%20the%20obese%22

Elliott, M. A., McGroarty, A., & Robertson, D. J. (2024). Using a volitional help sheet to increase university students' attendance at synchronous online lectures: A randomized controlled trial. *British Journal of Educational Psychology, 94*(2), 328–345. https://doi.org/10.1111/bjep.12647

Entringer, T. M., Gebauer, J. E., & Kröger, H. (2023). Big Five personality and religiosity: Bidirectional cross-lagged effects and their moderation by culture. *Journal of Personality, 91*(3), 736–752. https://doi.org/10.1111/jopy.12770

Epton, T., Currie, S., & Armitage, C. J. (2017). Unique effects of setting goals on behavior change: Systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology, 85*(12), 1182–1198. https://doi.org/10.1037/ccp0000260

European Values Study. (n.d.). *The TRAPD method for survey translation*. https://europeanvaluesstudy.eu/surveys/surveys-evs/evs-2017/methodology/the-trapd-method-for-survey-translation/

Evers, A., et al. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema, 25*(3), 283–291. https://doi.org/10.7334/psicothema2013.97

Facer-Childs, E., & Brandstaetter, R. (2015). The impact of circadian phenotype and time since awakening on diurnal performance in athletes. *Current Biology, 25*(4), 518–522. https://research.birmingham.ac.uk/en/publications/the-impact-of-circadian-phenotype-and-time-since-awakening-on-diu/

Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin, 136*(5), 894–914. https://doi.org/10.1037/a0019993

Feinberg, R. A., & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice, 35*(2), 36–49. https://ncme.org/wp-content/uploads/2025/10/Module-42-Simulation-Studies-in-Psychometrics-Fei.pdf

Fischer, D., Lombardi, D. A., Marucci-Wellman, H., & Roenneberg, T. (2017). Chronotypes in the US: Influence of age and sex. *PLoS ONE, 12*(6), e0178782. https://doi.org/10.1371/journal.pone.0178782

Fishbach, A., & Dhar, R. (2005). Goals as excuses or guides. *Journal of Consumer Research, 32*(3), 370–377. https://ideas.repec.org/a/oup/jconrs/v32y2005i3p370-377.html

Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology, 44*(1), 118–123. https://doi.org/10.1037/h0059240

Fuller, B. E., Holland, J. L., & Johnston, J. A. (1999). The relation of profile elevation in the Self-Directed Search to personality variables. *Journal of Career Assessment, 7*(2), 111–123. https://doi.org/10.1177/106907279900700202

Furnham, A. (1996). The big five versus the big four: The relationship between the MBTI and NEO-PI five factor model. *Personality and Individual Differences, 21*(2), 303–307. https://doi.org/10.1016/0191-8869(96)00033-5

Furnham, A., Moutafi, J., & Crump, J. (2003). The relationship between the revised NEO-Personality Inventory and the Myers-Briggs Type Indicator. *Social Behavior and Personality, 31*(6), 577–584. https://doi.org/10.2224/sbp.2003.31.6.577

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*(2), 349–360. https://academic.oup.com/poq/article-abstract/73/2/349/1832998

Gallagher, K. M., & Updegraff, J. A. (2012). Health message framing effects on attitudes, intentions, and behavior: A meta-analytic review. *Annals of Behavioral Medicine, 43*(1), 101–116. https://doi.org/10.1007/s12160-011-9308-7

Gallup. (n.d.). *CliftonStrengths (StrengthsFinder)*. https://www.gallup.com/cliftonstrengths/en/strengthsfinder.aspx

Gardner, W. L., & Martinko, M. J. (1996). Using the MBTI to study managers: A literature review and research agenda. *Journal of Management, 22*(1), 45–83. https://doi.org/10.1177/014920639602200103

Gerlach, G., Herpertz, S., & Loeber, S. (2015). Personality traits and obesity: A systematic review. *Obesity Reviews, 16*(1), 32–63. https://doi.org/10.1111/obr.12235

Gerlach, M., Farb, B., Revelle, W., & Amaral, L. A. N. (2018). A robust data-driven approach identifies four personality types across four large data sets. *Nature Human Behaviour, 2*, 735–742. https://doi.org/10.1038/s41562-018-0419-z

Gerlach, M., Revelle, W., & Amaral, L. A. N. (2019). Reply to: Four personality types may be neither robust nor exhaustive. *Nature Human Behaviour, 3*(10), 1047–1048. https://doi.org/10.1038/s41562-019-0722-3

Gherzi, S., Egan, D., Stewart, N., Haisley, E., & Ayton, P. (2014). The meerkat effect: Personality and market returns affect investors' portfolio monitoring behaviour. *Journal of Economic Behavior & Organization, 107*, 512–526. https://openaccess.city.ac.uk/id/eprint/16280/1/Meerkat_July2014_Final.pdf

Ghosh, A. P., & Huang, L. (2020). *Dynamic budget monitoring: When access to budget feedback leads to increase in spending* [CEPR technical report]. https://cepr.org/system/files/2022-08/Dynamic%20Budget%20Monitoring%20-%2

0When%20Access%20to%20Budget%20Feedback%20Leads%20to%20Increase%20in%20Spending%20-%20Anastasiya%20Pocheptsova%20Ghosh%20&%20Liang%20Huang.pdf

Glavin, K. W., & Savickas, M. L. (2011). Interpreting Self-Directed Search profiles: Validity of the "Rule of Eight". *Journal of Vocational Behavior, 79*(2), 414–418. https://doi.org/10.1016/j.jvb.2011.01.005

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics, 112*(2), 631–645. https://Rady.ucsd.edu/_files/faculty-research/uri-gneezy/risk-taking.pdf

Gold, J. M., et al. (2020). Conscientiousness in weight loss maintainers and regainers. *Health Psychology, 39*(5), 421–429. https://doi.org/10.1037/hea0000846

Goldberg, L. R., & Somer, O. (2000). The hierarchical structure of common Turkish person-descriptive adjectives. *European Journal of Personality, 14*(6), 497–531. https://doi.org/10.1002/1099-0984(200011/12)14:6%3C497::AID-PER379%3E3.0.CO;2-R

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology, 38*, 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1

Gottman, J. M., & Levenson, R. W. (2000). The timing of divorce: Predicting when a couple will divorce over a 14-year period. *Journal of Marriage and Family, 62*(3), 737–745. https://doi.org/10.1111/j.1741-3737.2000.00737.x

Goyal, M., et al. (2014). Meditation programs for psychological stress and well-being: A systematic review and meta-analysis. *JAMA Internal Medicine, 174*(3), 357–368. https://www.uned.es/universidad/inicio/dam/jcr:5a55d4bf-c8bf-4797-aca6-c0e85802188d/Goyal%20(2014)%20Meta-analisis%20mindfulness.pdf

Graham, E. K., et al. (2017). Personality predicts mortality risk: An integrative data analysis of 15 international longitudinal studies. *Journal of Research in Personality, 70*, 174–186. https://stacks.cdc.gov/view/cdc/53475/cdc_53475_DS1.pdf

Grant, A. M. (2013). Rethinking the extraverted sales ideal: The ambivert advantage. *Psychological Science, 24*(6), 1024–1030. https://doi.org/10.1177/0956797612463706

Gregory, C., & Lewis, P. (2016). *Linking client assessment profiles to O*NET occupational profiles within the O*NET Interest Profiler Short Form and Mini Interest Profiler (Mini-IP)*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/Mini-IP_Linking.pdf

Grüner, H. P., & Rohde, K. (2022). *The value of choice* (CEPR Discussion Paper No. 17425). https://ideas.repec.org/p/cpr/ceprdp/17425.html

Hakulinen, C., et al. (2015). Personality and smoking: Individual-participant meta-analysis of nine cohort studies. *Addiction, 110*(11), 1844–1852. https://pmc.ncbi.nlm.nih.gov/articles/PMC4609271

Halko, S., & Kientz, J. A. (2010). Personality and persuasive technology: An exploratory study on health-promoting mobile applications. *Persuasive 2010, Lecture Notes in Computer Science*, 150–161. https://faculty.washington.edu/jkientz/papers/Halko-Personality-Persuasive2010.pdf

Hammond, M. M., Neff, N. L., Farr, J. L., Schwall, A. R., & Zhao, X. (2011). Predictors of individual-level innovation at work: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts, 5*(1), 90–105. https://doi.org/10.1037/a0018556

Hansen, J.-I. C., Dik, B. J., & Zhou, S. (2008). An examination of the structure of leisure interests of college students, working-age adults, and retirees. *Journal of Counseling Psychology, 55*(2), 133–145. https://doi.org/10.1037/0022-0167.55.2.133

Harkin, B., et al. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin, 142*(2), 198–229. https://doi.org/10.1037/bul0000025

Harris, K., & Vazire, S. (2016). On friendship development and the Big Five personality traits. *Social and Personality Psychology Compass, 10*(11), 647–667. https://doi.org/10.1111/spc3.12287

Heller, D., Watson, D., & Ilies, R. (2004). The role of person versus situation in life satisfaction: A critical examination. *Psychological Bulletin, 130*(4), 574–600. https://doi.org/10.1037/0033-2909.130.4.574

HEXACO. (n.d.). *The HEXACO Personality Inventory – Revised*. https://hexaco.org/hexaco-inventory

Hirsh, J. B., & Inzlicht, M. (2008). The devil you know: Neuroticism predicts neural response to uncertainty. *Psychological Science, 19*(10), 962–967. https://www.psychologicalscience.org/news/releases/uncertainty-can-be-more-stressful-than-clear-negative-feedback.html

Hirsh, J. B., Kang, S. K., & Bodenhausen, G. V. (2012). Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits. *Psychological Science, 23*(6), 578–581. https://doi.org/10.1177/0956797611436349

Hoeglund, T. J., & Hansen, J.-I. C. (1999). Holland-style measures of congruence: Are complex indices more effective predictors of satisfaction? *Journal of Vocational Behavior, 54*(3), 471–482. https://doi.org/10.1006/jvbe.1998.1675

Hoerger, M. (2010). Participant dropout as a function of survey length in Internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior, and Social Networking, 13*(6), 697–700. https://doi.org/10.1089/cyber.2009.0445

Hoff, K. A., Briley, D. A., Wee, C. J. M., & Rounds, J. (2018). Normative changes in interests from adolescence to adulthood: A meta-analysis of longitudinal studies. *Psychological Bulletin, 144*(4), 426–451. https://www.pubmed.ncbi.nlm.nih.gov/29494193/

Hook, J. N., Hall, T. W., Davis, D. E., Van Tongeren, D. R., & Conner, M. (2021). The Enneagram: A systematic review of the literature and directions for future research. *Journal of Clinical Psychology, 77*(4), 865–883. https://pubmed.ncbi.nlm.nih.gov/33332604/

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4*(2), 97–110. https://pubmed.ncbi.nlm.nih.gov/1027738/

Hřebíčková, M., Jelínek, M., Květon, P., Benkovič, A., Botek, M., Sudzina, F., Soto, C. J., & John, O. P. (2020). Big Five Inventory 2 (BFI-2): Hierarchický model s 15 subškálami. *Československá psychologie, 64*(4), 437–460. https://webarchiv.lib.cas.cz/wayback/20201006133032/http:/cspsych.psu.cas.cz/result.php?id=1123

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Humbad, M. N., Donnellan, M. B., Iacono, W. G., McGue, M., & Burt, S. A. (2010). Is spousal similarity for personality a matter of convergence or selection? *Personality and Individual Differences, 49*(7), 827–830. https://doi.org/10.1016/j.paid.2010.07.010

Iachan, R. (1984). A measure of agreement for use with the Holland classification system. *Journal of Vocational Behavior, 24*(2), 133–141. https://doi.org/10.1016/0001-8791(84)90001-0

ING Türkiye. (2023). *Tasarruf Eğilimleri Araştırması 2023 4. çeyrek* (IPSOS) [Press report]. https://www.sozcu.com.tr/tasarruf-sahiplerinin-tercihi-altin-ve-nakit-oldu-p24545 (reference to be verified)

International Test Commission. (2005). *International guidelines on computer-based and internet-delivered testing*. https://www.intestcom.org/files/guideline_computer_based_testing.pdf

International Test Commission. (2017). The ITC guidelines for translating and adapting tests (2nd ed.). *International Journal of Testing, 18*(2), 101–134. https://doi.org/10.1080/15305058.2017.1398166

IPIP. (n.d.-a). *Permission*. International Personality Item Pool. https://ipip.ori.org/newPermission.htm

IPIP. (n.d.-b). *Translations*. International Personality Item Pool. https://ipip.ori.org/newTranslations.htm

IPIP. (n.d.-c). *Key to the NEO facets*. International Personality Item Pool. https://ipip.ori.org/newNEO_FacetsKey.htm

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy, 3*(2), 159–186. https://www.cambridge.org/core/journals/behavioural-public-policy/article/when-and-why-defaults-influence-decisions-a-metaanalysis-of-default-effects/67AF6972CFB52698A60B6BD94B70C2C0

Jacques-Hamilton, R., Sun, J., & Smillie, L. D. (2019). Costs and benefits of acting extraverted: A randomized controlled trial. *Journal of Experimental Psychology: General, 148*(9), 1538–1556. https://doi.org/10.1037/xge0000516

Jensen-Campbell, L. A., & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality, 69*(2), 323–362. https://doi.org/10.1111/1467-6494.00148

Jeronimus, B. F., Kotov, R., Riese, H., & Ormel, J. (2016). Neuroticism's prospective association with mental disorders halves after adjustment for baseline symptoms and psychiatric history. *Psychological Medicine, 46*(14), 2883–2906. https://www.cambridge.org/core/journals/psychological-medicine/article/neuroticisms-prospective-association-with-mental-disorders-halves-after-adjustment-for-baseline-symptoms-and-psychiatric-history-but-the-adjusted-association-hardly-decays-with-time-a-metaanalysis-on-59-longitudinalprospective-studies-with-443-313-participants/35DF609ABE724C4F25C247D03C38BB3C

Jia, Y., Xu, B., Karanam, Y., & Voida, S. (2016). Personality-targeted gamification: A survey study on personality traits and motivational affordances. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2001–2013. https://doi.org/10.1145/2858036.2858515

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Johnson, J. A. (n.d.). *IPIP-NEO-120 item-level data* [Data set: IPIP120.dat]. Open Science Framework. https://osf.io/wxvth/

Johnson, J. A. (2020). [Calibrating personality self-report scores to acquaintance ratings]. https://bpb-us-e1.wpmucdn.com/sites.psu.edu/dist/1/163537/files/2023/07/Calibrating2020.pdf (reference to be verified)

Jokela, M., et al. (2013a). Association of personality with the development and persistence of obesity: A meta-analysis based on individual-participant data. *Obesity Reviews, 14*(4), 315–323. https://doi.org/10.1111/obr.12007

Jokela, M., et al. (2013b). Personality and all-cause mortality: Individual-participant meta-analysis of 3,947 deaths in 76,150 adults. *American Journal of Epidemiology, 178*(5), 667–675. https://pmc.ncbi.nlm.nih.gov/articles/PMC3755650

Joshanloo, M. (2022). Personality trait level and change predict future financial well-being. *Personality and Individual Differences, 191*, 111575. https://flosse.dss.gov.au/flossejspui/handle/10620/18954

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86*(1), 80–92. https://doi.org/10.1037/0021-9010.86.1.80

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*(4), 765–780. https://doi.org/10.1037/0021-9010.87.4.765

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87*(3), 530–541. https://doi.org/10.1037/0021-9010.87.3.530

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*(3), 621–652. https://doi.org/10.1111/j.1744-6570.1999.tb00174.x

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87*(4), 797–807. https://doi.org/10.1037/0021-9010.87.4.797

Judge, T. A., Livingston, B. A., & Hurst, C. (2012). Do nice guys—and gals—really finish last? The joint effects of sex and agreeableness on income. *Journal of Personality and Social Psychology, 102*(2), 390–407. https://doi.org/10.1037/a0026021

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance. *Journal of Applied Psychology, 98*(6), 875–925. https://doi.org/10.1037/a0033901

Kajonius, P. J., & Johnson, J. A. (2019). Assessing the structure of the Five Factor Model of Personality (IPIP-NEO-120) in the public domain. *Europe's Journal of Psychology, 15*(2), 260–275. https://pmc.ncbi.nlm.nih.gov/articles/PMC7871748/

Kappes, A., Oettingen, G., & Pak, H. (2012). Mental contrasting and the self-regulation of responding to negative feedback. *Personality and Social Psychology Bulletin, 38*(7), 845–857. https://doi.org/10.1177/0146167212446833

Kariyer Planım. (n.d.). *Kariyer profil testi*. http://www.kariyerplanim.com/

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science, 62*(12), 3393–3411. https://doi.org/10.1287/mnsc.2015.2296

Kashdan, T. B., et al. (2018). The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality, 73*, 130–149. https://doi.org/10.1016/j.jrp.2017.11.011

Katahira, K., Kunisato, Y., Yamashita, Y., & Suzuki, S. (2020). Commentary: A robust data-driven approach identifies four personality types across four large data sets. *Frontiers in Big Data, 3*, 8. https://doi.org/10.3389/fdata.2020.00008

Keirsey, D. (n.d.). *Temperament overview*. https://keirsey.com/temperament-overview/

Keller, C., & Siegrist, M. (2015). Does personality influence eating styles and food choices? Direct and indirect effects. *Appetite, 84*, 128–138. https://europepmc.org/article/MED/25308432

Keller, J., et al. (2021). Habit formation following routine-based versus time-based cue planning: A randomized controlled trial. *British Journal of Health Psychology, 26*(3), 807–824. https://doi.org/10.1111/bjhp.12504

Kelly, K. R., & Jugovic, H. (2001). Concurrent validity of the online version of the Keirsey Temperament Sorter II. *Journal of Career Assessment, 9*(1), 49–59. https://doi.org/10.1177/106907270100900104

Kern, M. L., & Friedman, H. S. (2008). Do conscientious individuals live longer? A quantitative review. *Health Psychology, 27*(5), 505–512. https://www.gwern.net/doc/psychology/personality/conscientiousness/2008-kern.pdf

Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The goal-gradient hypothesis resurrected. *Journal of Marketing Research, 43*(1), 39–58. https://home.uchicago.edu/ourminsky/Goal-Gradient_Illusionary_Goal_Progress.pdf

Kliger, D., & Gilad, D. (2012). Red light, green light: Color priming in financial decisions. *Journal of Socio-Economics, 41*(5), 738–745. https://ideas.repec.org/a/eee/soceco/v41y2012i5p738-745.html

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. *Psychological Bulletin, 119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Kong, S. S. (2010). Relationships between Myers-Briggs Type Indicator (MBTI) psychological type and marital satisfaction, divorce proneness, positive affect, and conflict regulation in clinic couples. *Journal of Korean Academy of Nursing, 40*(3), 336–348. https://doi.org/10.4040/jkan.2010.40.3.336

Koo, M., & Fishbach, A. (2012). The small-area hypothesis: Effects of progress monitoring on goal adherence. *Journal of Consumer Research, 39*(3), 493–509. https://academic.oup.com/jcr/article/39/3/493/1795380

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin, 136*(5), 768–821. https://researchconnect.stonybrook.edu/en/publications/linking-big-personality-traits-to-anxiety-depressive-and-substanc/

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit. *Personnel Psychology, 58*(2), 281–342. https://doi.org/10.1111/j.1744-6570.2005.00672.x

Kung, F. Y. H., Kwok, N., & Brown, D. J. (2018). Are attention check questions a threat to scale validity? *Applied Psychology, 67*(2), 264–283. https://doi.org/10.1111/apps.12108

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist, 64*(4), 241–256. https://pmc.ncbi.nlm.nih.gov/articles/PMC2792076/

Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology, 40*(6), 998–1009. https://doi.org/10.1002/ejsp.674

Lambert, N. M., Clark, M. S., Durtschi, J., Fincham, F. D., & Graham, S. M. (2010). Benefits of expressing gratitude: Expressing gratitude to a partner changes one's view of the relationship. *Psychological Science, 21*(4), 574–580. https://doi.org/10.1177/0956797610364003

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*(2), 202–220. https://doi.org/10.1177/1094428105284919

Lavigne, K. M., Hofman, S., Ring, A. J., Ryder, A. G., & Woodward, T. S. (2013). The personality of meaning in life. *Journal of Positive Psychology, 8*(1), 34–43. https://instituteofcoaching.org/node/34290

Lenhard, W., & Lenhard, A. (2021). Improvement of norm score quality via regression-based continuous norming. *Educational and Psychological Measurement, 81*(2), 229–261. https://doi.org/10.1177/0013164420928457

Letkiewicz, J. C., & Fox, J. J. (2014). Conscientiousness, financial literacy, and asset accumulation of young adults. *Journal of Consumer Affairs, 48*(2), 274–300. https://doi.org/10.1111/joca.12040

Lipnevich, A. A., Credé, M., Hahn, E., Spinath, F. M., Roberts, R. D., & Preckel, F. (2017). How distinctive are morningness and eveningness from the Big Five factors of personality? A meta-analytic investigation. *Journal of Personality and Social Psychology, 112*(3), 491–509. https://www.ncbi.nlm.nih.gov/pubmed/27977220

Liu, J., Hoff, K. A., Chu, C., Oswald, F. L., & Rounds, J. (2025). Toward whole-person fit assessment: Integrating interests, values, skills, knowledge, and personality using O*NET. *Journal of Applied Psychology, 110*(5), 623–647. https://doi.org/10.1037/apl0001232

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57–64. https://econtent.hogrefe.com/doi/10.1027/1614-2241.2.2.57

Low, K. S. D., Yoon, M., Roberts, B. W., & Rounds, J. (2005). The stability of vocational interests from early adolescence to middle adulthood: A quantitative review of longitudinal studies. *Psychological Bulletin, 131*(5), 713–737. https://doi.org/10.1037/0033-2909.131.5.713

Lukas, M., & Howard, R. C. (2023). The influence of budgets on consumer spending. *Journal of Consumer Research, 49*(5), 697–720. https://academic.oup.com/jcr/article/49/5/697/6603733

Lundmark, S., Silber, H., Rossmann, J., & Gummer, T. (2023). *Comparing types of attentiveness measures*. University of Gothenburg. https://www.gu.se/sites/default/files/2023-09/Comparing%20types%20of%20attentiveness%20measures%20-%20Lundmark%20Silber%20Rossman%20%20Gummer%20%282023%29.pdf (reference to be verified)

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19–40. https://psychology.psy.sunysb.edu/attachment/measures/content/maccallum_on_dichotomizing.pdf

Malouff, J. M., Thorsteinsson, E. B., Rooke, S. E., & Schutte, N. S. (2007). Alcohol involvement and the Five-Factor Model of personality: A meta-analysis. *Journal of Drug Education, 37*(3), 277–294. https://rune.une.edu.au/web/handle/1959.11/4125

Malouff, J. M., Thorsteinsson, E. B., Schutte, N. S., Bhullar, N., & Rooke, S. E. (2010). The Five-Factor Model of personality and relationship satisfaction of intimate partners: A meta-analysis. *Journal of Research in Personality, 44*(1), 124–127. https://doi.org/10.1016/j.jrp.2009.09.004

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*(3), 471–491. https://doi.org/10.1037/a0019227

Martínez, G. J., et al. (2021). Predicting participant compliance with fitness tracker wearing and ecological momentary assessment protocols in information workers: Observational study. *JMIR mHealth and uHealth, 9*(11), e22218. https://doi.org/10.2196/22218

Masi, C. M., Chen, H. Y., Hawkley, L. C., & Cacioppo, J. T. (2011). A meta-analysis of interventions to reduce loneliness. *Personality and Social Psychology Review, 15*(3), 219–266. https://doi.org/10.1177/1088868310377394

Matz, S. C., Gladstone, J. J., & Stillwell, D. (2016). Money buys happiness when spending fits our personality. *Psychological Science, 27*(5), 715–725. https://doi.org/10.1177/0956797616635200

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences, 114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

Mavletova, A., & Couper, M. P. (2014). [Breakoff rates in mobile and desktop web surveys; GOR 2014 conference presentation]. https://cts.hse.ru/data/2014/04/03/1317876257/Mavletova_Couper_GOR_2014.pdf (reference to be verified)

Mavletova, A., & Couper, M. P. (2015). A meta-analysis of breakoff rates in mobile web surveys. In D. Toninelli, R. Pinter, & P. de Pedraza (Eds.), *Mobile research methods*. Ubiquity Press. https://dx.doi.org/10.5334/bar.f (reference to be verified)

Mazri, F. H., et al. (2020). The association between chronotype and dietary pattern among adults: A scoping review. *International Journal of Environmental Research and Public Health, 17*(1), 68. https://doi.org/10.3390/ijerph17010068

McCloy, R., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999a). *Development of the O*NET paper-and-pencil Work Importance Locator*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/DevWIL.pdf

McCloy, R., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999b). *Development of the O*NET computerized Work Importance Profiler*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/DevCWIP.pdf

McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality, 57*(1), 17–40. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x

McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 70*(3), 552–566. https://doi.org/10.1037/0022-3514.70.3.552

McMahan, E. A., & Estes, D. (2015). The effect of contact with natural environments on positive and negative affect: A meta-analysis. *Journal of Positive Psychology, 10*(6), 507–519. https://instituteofcoaching.org/node/35082

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Molloy, G. J., O'Carroll, R. E., & Ferguson, E. (2014). Conscientiousness and medication adherence: A meta-analysis. *Annals of Behavioral Medicine, 47*(1), 92–101. https://doi.org/10.1007/s12160-013-9524-4

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships, 25*(6), 889–922. https://doi.org/10.1177/0265407508096700

Morillo, D., Abad, F. J., Kreitchmann, R. S., Leenen, I., Hontangas, P., & Ponsoda, V. (2019). The journey from Likert to forced-choice questionnaires: Evidence of the invariance of item parameters. *Revista de Psicología del Trabajo y de las Organizaciones, 35*(2), 75–83. https://doi.org/10.5093/jwop2019a11

Morrison, L. G., et al. (2017). The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: An exploratory trial. *PLoS ONE, 12*(1), e0169162. https://doi.org/10.1371/journal.pone.0169162

Morsünbül, Ü. (2014). [Turkish adaptation of the Quick Big Five Personality Test: validity and reliability]. *Düşünen Adam, 27*, 316–322. https://dusunenadamdergisi.org/storage/upload/pdfs/1585750117-en.pdf (reference to be verified)

Mosca, I., & McCrory, C. (2016). Personality and wealth accumulation among older couples: Do dispositional characteristics pay dividends? *Journal of Economic Psychology, 56*, 1–19. https://mural.maynoothuniversity.ie/id/eprint/11388/1/IM_Personality_2016.pdf

Mund, M., & Neyer, F. J. (2014). Treating personality–relationship transactions with respect: Narrow facets, advanced models, and extended time frames. *Journal of Personality and Social Psychology, 107*(2), 352–368. https://doi.org/10.1037/a0036719

Munro, I. A., Bore, M. R., Munro, D., & Garg, M. L. (2011). Using personality as a predictor of diet induced weight loss and weight management. *International Journal of Behavioral Nutrition and Physical Activity, 8*, 129. https://europepmc.org/article/MED/22112231

National Center for O*NET Development. (n.d.-a). *O*NET Interest Profiler Short Form psychometric characteristics*. https://www.onetcenter.org/dl_files/IPSF_Psychometric.pdf

National Center for O*NET Development. (n.d.-b). *O*NET Interest Profiler manual*. https://www.onetcenter.org/dl_files/IP_Manual.pdf

National Center for O*NET Development. (n.d.-c). *Work Importance Locator: Instructions*. https://www.onetcenter.org/dl_files/WIL_zips/WIL-Instr-deskv.pdf

National Center for O*NET Development. (n.d.-d). *O*NET Tools Developer License*. https://www.onetcenter.org/license_toolsdev.html

Ndulue, C., Oyebode, O., Iyer, R. S., Ganesh, A., Ahmed, S. I., & Orji, R. (2022). Personality-targeted persuasive gamified systems: Exploring the impact of application domain on the effectiveness of behaviour change strategies. *User Modeling and User-Adapted Interaction, 32*(1–2), 165–214. https://doi.org/10.1007/s11257-022-09319-w

Neyer, F. J., & Lehnart, J. (2007). Relationships matter in personality development: Evidence from an 8-year longitudinal study across young adulthood. *Journal of Personality, 75*(3), 535–568. https://doi.org/10.1111/j.1467-6494.2007.00448.x

Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology, 58*(2), 367–408. https://doi.org/10.1111/j.1744-6570.2005.00515.x

Noftle, E. E., & Shaver, P. R. (2006). Attachment dimensions and the Big Five personality traits: Associations and comparative ability to predict relationship quality. *Journal of Research in Personality, 40*(2), 179–208. http

s://doi.org/10.1016/j.jrp.2004.11.003

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology, 22*(3), 453–460. https://doi.org/10.1016/j.jcps.2011.08.002

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science, 7*(4), 384–403. https://europepmc.org/abstract/MED/26168474

Oehler, A., Wendt, S., Wedlich, F., & Horn, M. (2018). Investors' personality influences investment decisions: Experimental evidence on extraversion and neuroticism. *Journal of Behavioral Finance, 19*(1), 30–48. https://doi.org/10.1080/15427560.2017.1366495

Oklahoma Baptist University. (2023). [Perceived accuracy of general and specific test results, Scholars Day poster]. https://scholarlycommons.obu.edu/scholars_day_conference/2023/posters/38 (reference to be verified)

Olafsson, A., & Pagel, M. (2017). *The ostrich in us: Selective attention to financial accounts, income, spending, and liquidity* (NBER Working Paper No. 23945). https://www.nber.org/papers/w23945

O*NET Resource Center. (n.d.-a). *O*NET Career Exploration Tools license*. https://www.onetcenter.org/license_tools.html

O*NET Resource Center. (n.d.-b). *O*NET Tools Developer License*. https://www.onetcenter.org/license_toolsdev.html

O*NET Resource Center. (n.d.-c). *O*NET Work Importance Profiler and Work Importance Locator*. https://www.onetcenter.org/WIP.html

O*NET Resource Center. (n.d.-d). *O*NET Database license*. https://www.onetcenter.org/license_db.html

Oosterhuis, H. E. M., van der Ark, L. A., & Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment, 23*(2), 191–202. https://doi.org/10.1177/1073191115580638

Open Heart. (2026). *Chronotype-aligned exercise timing* [Randomised controlled trial; ISRCTN14276441]. https://bmjgroup.com/timing-exercise-to-match-body-clock-chronotype-may-lower-cardiovascular-disease-risk/ (reference to be verified)

Oshio, A., Taku, K., Hirano, M., & Saeed, G. (2018). Resilience and Big Five personality traits: A meta-analysis. *Personality and Individual Differences, 127*, 54–60. https://doi.org/10.1016/j.paid.2018.01.048

Oyibo, K., Orji, R., & Vassileva, J. (2017). Investigation of the influence of personality traits on Cialdini's persuasive strategies. *Personalization in Persuasive Technology Workshop, CEUR Workshop Proceedings, 1833*. http://ceur-ws.org/Vol-1833/4_Oyibo.pdf

Pacanowski, C. R., Linde, J. A., & Neumark-Sztainer, D. (2015). Self-weighing: Helpful or harmful for psychological well-being? *Current Obesity Reports, 4*(1), 65–72. https://europepmc.org/article/MED/26627092

PAR. (n.d.). *Self-Directed Search handbook, Chapter 1*. Psychological Assessment Resources. https://www.parinc.com/Portals/0/Webuploads/samplerpts/SDS%20Handbook%20Chapter%201.pdf

Parks-Leduc, L., Feldman, G., & Bardi, A. (2015). Personality traits and personal values: A meta-analysis. *Personality and Social Psychology Review, 19*(1), 3–29. https://pure.royalholloway.ac.uk/en/publications/personality-traits-and-personal-values-a-meta-analysis/

Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*(3), 524–539. https://doi.org/10.1037/0022-3514.81.3.524

Perkmen, S., & Tezci, E. (2015). [A 55-item Holland-based interest scale in pre-service teachers]. Balıkesir University open access. https://acikerisim.balikesir.edu.tr/items/d7bef6e3-02d7-4e88-91b1-38e45092a012/full (reference to be verified)

Pfattheicher, S., et al. (2017). Old wine in new bottles? The case of self-compassion and neuroticism. *European Journal of Personality, 31*(2), 160–169. https://doi.org/10.1002/per.2097

Phillips, L. A., Burns, E., & Leventhal, H. (2021). Time-of-day differences in treatment-related habit strength and adherence. *Annals of Behavioral Medicine, 55*(3), 280–285. https://doi.org/10.1093/abm/kaaa042

Pittenger, D. J. (1993). The utility of the Myers-Briggs Type Indicator. *Review of Educational Research, 63*(4), 467–488. https://doi.org/10.3102/00346543063004467

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research, 57*(3), 210–221. https://doi.org/10.1037/1065-9293.57.3.210

Poškus, M. S., Kairys, A., Liniauskaitė, A., & Žukauskienė, R. (2014). "Your NEO summary": Is it a valid way of providing feedback, or is its accuracy due to the Barnum effect? *Psichologija, 49*, 44–59. https://doi.org/10.15388/Psichol.2014.49.3694

Primi, R., De Fruyt, F., Santos, D., Antonoplis, S., & John, O. P. (2020). True or false? Keying direction and acquiescence influence the validity of socio-emotional skills items in predicting high school achievement. *International Journal of Testing, 20*(2), 97–121. https://doi.org/10.1080/15305058.2019.1673398

Prinzie, P., Stams, G. J. J. M., Deković, M., Reijntjes, A. H. A., & Belsky, J. (2009). The relations between parents' Big Five personality factors and parenting: A meta-analytic review. *Journal of Personality and Social Psychology, 97*(2), 351–362. https://doi.org/10.1037/a0015823

Pullmann, H., Raudsepp, L., & Allik, J. (2006). Stability and change in adolescents' personality: A longitudinal study. *European Journal of Personality, 20*(6), 447–459. https://doi.org/10.1002/per.611

Pündük, Z., Gür, H., & Ercan, İ. (2005). Sabahçıl-akşamcıl anketi Türkçe uyarlamasında güvenilirlik çalışması. *Türk Psikiyatri Dergisi, 16*(1), 40–45. https://avesis.uludag.edu.tr/yayin/99c6a668-bc64-4417-bd78-6a3a55b1d31c/a-reliability-study-of-the-turkish-version-of-the-mornings-evenings-questionnaire-sabahcil-aksamcil-anketi-turkce-uyarlamasinda-guvenilirlik-calismasi

Purcell, K., et al. (2014). The effect of rate of weight loss on long-term weight management: A randomised controlled trial. *The Lancet Diabetes & Endocrinology, 2*(12), 954–962. https://doi.org/10.1016/S2213-8587(14)70200-1

Purdue University. (n.d.). [The power of expressed and measured interests to predict current choice; doctoral dissertation]. https://docs.lib.purdue.edu/dissertations/AAI8400378 (reference to be verified)

Putka, D. J., Kell, H. J., Voss, N., Oswald, F. L., & Lewis, P. (2024). *Revisiting the O*NET Work Styles domain*. National Center for O*NET Development. https://onetcenter.org/dl_files/Work_Styles_New.pdf

Rammstedt, B., Kemper, C. J., & Borg, I. (2013). Correcting Big Five personality measurements for acquiescence: An 18-country cross-cultural study. *European Journal of Personality, 27*(1), 71–81. https://doi.org/10.1002/per.1894

Remetesting. (n.d.). *Test satış*. https://remetesting.com/Home/TestSatis

Revilla, M., & Höhne, J. K. (2020). How long do respondents think online surveys should be? New evidence from two online panels in Germany. *International Journal of Market Research, 62*(5), 538–545. https://www.sowi.uni-mannheim.de/media/Einrichtungen/gip/Bilder/Dokumente/revilla_hoehne__2020__-_online_survey_length.pdf

Rhodes, R. E., & Smith, N. E. I. (2006). Personality correlates of physical activity: A review and meta-analysis. *British Journal of Sports Medicine, 40*, 958–965. https://bjsm.bmj.com/content/40/12/958

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality. *Perspectives on Psychological Science, 2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin, 143*(2), 117–141. https://doi.org/10.1037/bul0000088

Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five personality factors and personal values. *Personality and Social Psychology Bulletin, 28*(6), 789–801. https://journals.sagepub.com/doi/10.1177/0146167202289008

Roenneberg, T., Pilz, L. K., Zerbini, G., & Winnebeck, E. C. (2019). Chronotype and social jetlag: A (self-) critical review. *Biology, 8*(3), 54. https://europepmc.org/article/PMC/PMC6784249

rOpenSci. (n.d.). *mctq: Tools to process the Munich ChronoType Questionnaire*. https://packages.ropensci.org/mctq/doc/manual.html

Rose, P., & Segrist, D. J. (2014). Negative and positive urgency may both be risk factors for compulsive buying. *Journal of Behavioral Addictions, 3*(2), 128–132. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4117284/

Ross, C. F. J., & Francis, L. J. (2010). The relationship of intrinsic, extrinsic, and quest religious orientations to Jungian psychological type among churchgoers in England and Wales. *Mental Health, Religion & Culture, 13*(7–8), 805–819. https://doi.org/10.1080/13674670802207462

Rounds, J., Wee, C. J. M., Cao, M., Song, C., & Lewis, P. (2016). *Development of an O*NET Mini Interest Profiler (Mini-IP) for mobile devices: Psychometric characteristics*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/Mini-IP.pdf

Rounds, J., & Su, R. (2014). The nature and power of interests. *Current Directions in Psychological Science, 23*(2), 98–103. https://doi.org/10.1177/0963721414522812

Rounds, J., Su, R., Lewis, P., & Rivkin, D. (2010). *O*NET Interest Profiler Short Form psychometric characteristics: Summary*. National Center for O*NET Development. https://www.onetcenter.org/dl_files/IPSF_Psychometric.pdf

Rudolph, C. W., Lavigne, K. N., & Zacher, H. (2017). Career adaptability: A meta-analysis. *Journal of Vocational Behavior, 98*, 17–34. https://doi.org/10.1016/j.jvb.2016.09.002

Rustichini, A., DeYoung, C. G., Anderson, J. E., & Burks, S. V. (2016). Toward the integration of personality theory and decision theory in explaining economic behavior: An experimental investigation. *Journal of Behavioral and Experimental Economics, 64*, 122–137. https://doi.org/10.1016/j.socec.2016.04.019 (secondary citation)

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection. *Journal of Applied Psychology, 107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Saroglou, V. (2002). Religion and the five factors of personality: A meta-analytic review. *Personality and Individual Differences, 32*(1), 15–25. https://www.psyreli.org/_files/ugd/ba70ca_9734c4f5bea2409dbf9fe0f286621d16.pdf

Saroglou, V. (2010). Religiousness as a cultural adaptation of basic traits: A Five-Factor Model perspective. *Personality and Social Psychology Review, 14*(1), 108–125. https://research.dial.uclouvain.be/entities/publication/8eb1e87d-7b39-4740-a9b1-f558fca4236d

Saucier, G. (2019). Is religiousness a form of variation in personality, or in culture, or neither? Conceptual issues and empirical indications. *Archive for the Psychology of Religion, 41*(3), 216–223. https://doi.org/10.1177/0084672419894677

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology, 38*(2), 173–212. https://www.um.edu.mt/library/oar/bitstream/123456789/21679/1/The%20geographic%20distribution%20of%20Big%20Five%20personality%20traits%20Patterns%20and%20profiles%20of%20human%20self-description%20across%2056%20nations.pdf

Shapiro, G. K., & Burchell, B. J. (2012). Measuring financial anxiety. *Journal of Neuroscience, Psychology, and Economics, 5*(2), 92–103. https://crr.bc.edu/avoidance-comes-with-financial-anxiety/

Sharp, B., Danenberg, N., & Bellman, S. (2018). Psychological targeting [Letter to the editor]. *Proceedings of the National Academy of Sciences, 115*(34), E7890. https://doi.org/10.1073/pnas.1810436115

Sheeran, P., Listrom, O., & Gollwitzer, P. M. (2024). The when and how of planning: Meta-analysis of the scope and components of implementation intentions in 642 tests. *European Review of Social Psychology*. https://doi.org/10.1080/10463283.2024.2334563

Sicherman, N., Loewenstein, G., Seppi, D. J., & Utkus, S. P. (2016). Financial attention. *Review of Financial Studies, 29*(4), 863–897. https://doi.org/10.1093/rfs/hhv073

Silverman, J., & Barasch, A. (2023). On or off track: How (broken) streaks affect consumer decisions. *Journal of Consumer Research, 49*(6), 1095–1117. https://doi.org/10.1093/jcr/ucac029

Singh, B., Murphy, A., Maher, C., & Smith, A. E. (2024). Time to form a habit: A systematic review and meta-analysis of health behaviour habit formation and its determinants. *Healthcare, 12*(23), 2488. https://doi.org/10.3390/healthcare12232488

Sirois, F. M., Kitner, R., & Hirsch, J. K. (2015). Self-compassion, affect, and health-promoting behaviors. *Health Psychology, 34*(6), 661–669. https://pubmed.ncbi.nlm.nih.gov/25243717/

Solomon, B. C., & Jackson, J. J. (2014). Why do personality traits predict divorce? Multiple pathways through satisfaction. *Journal of Personality and Social Psychology, 106*(6), 978–996. https://doi.org/10.1037/a0036190

Somer, O., Korkmaz, M., & Tatar, A. (2002). Beş Faktör Kişilik Envanteri'nin geliştirilmesi-I: Ölçek ve alt ölçeklerin oluşturulması. *Türk Psikoloji Dergisi, 17*(50), 21–36. https://avesis.mcbu.edu.tr/yayin/9f580945-7e2e-48e8-9a28-d68cfc6c173e/development-of-five-factor-personality-inventory-bes-faktor-kisilik-envanterinin-gelistirilmesi-i-olcek-ve-alt-olceklerin-olusturulmasi

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143. https://pubmed.ncbi.nlm.nih.gov/27055049/

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of Big Five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718–737. https://doi.org/10.1037/0022-3514.94.4.718

Stach, M., Mulansky, L., Reichert, M., Pryss, R., & Beierle, F. (2024). Call to action: Investigating interaction delay in smartphone notifications. *Sensors, 24*(8), 2612. https://europepmc.org/article/PMC/PMC11053777

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240–261. https://doi.org/10.1037/0033-2909.124.2.240

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin, 133*(1), 65–94. https://doi.org/10.1037/0033-2909.133.1.65

Steger, M. F., Kashdan, T. B., Sullivan, B. A., & Lorentz, D. (2008). Understanding the search for meaning in life. *Journal of Personality, 76*(2), 199–228. https://mason.gmu.edu/~tkashdan/publications/searchmeaning_JP.pdf

Stein, R., & Swan, A. B. (2019). Evaluating the validity of Myers-Briggs Type Indicator theory: A teaching tool and window into intuitive psychology. *Social and Personality Psychology Compass, 13*(2), e12434. https://doi.org/10.1111/spc3.12434

Stephan, Y., Sutin, A. R., Bayard, S., Križan, Z., & Terracciano, A. (2018). Personality and sleep quality: Evidence from four prospective studies. *Health Psychology, 37*(3), 271–281. https://pmc.ncbi.nlm.nih.gov/articles/PMC5837948

Stieger, M., Flückiger, C., Rüegger, D., Kowatsch, T., Roberts, B. W., & Allemand, M. (2021). Changing personality traits with the help of a digital personality change intervention. *Proceedings of the National Academy of Sciences, 118*(8), e2017548118. https://doi.org/10.1073/pnas.2017548118

Stoll, G., Rieger, S., Lüdtke, O., Nagengast, B., Trautwein, U., & Roberts, B. W. (2017). Vocational interests assessed at the end of high school predict life outcomes assessed 10 years later over and above IQ and Big Five personality traits. *Journal of Personality and Social Psychology, 113*(1), 167–184. https://europepmc.org/search?q

uery=TITLE%3A%22Vocational%20interests%20assessed%20at%20the%20end%20of%20high%20school%22

Strickhouser, J. E., Zell, E., & Krizan, Z. (2017). Does personality predict health and well-being? A metasynthesis. *Health Psychology, 36*(8), 797–810. https://europepmc.org/article/MED/28277701

Studia Doctoralia. (n.d.). [Visual format of IPIP feedback and perceived accuracy]. https://journals.unibuc.ro/index.php/sd/article/view/1300 (reference to be verified)

Süerdem, A., & Erkök, B. (2016). Assessing the reliability and validity of a shorter version of RIASEC in Turkish. *SHS Web of Conferences, 26*, 01063. https://doi.org/10.1051/shsconf/20162601063

Sutin, A. R., Ferrucci, L., Zonderman, A. B., & Terracciano, A. (2011). Personality and obesity across the adult life span. *Journal of Personality and Social Psychology, 101*(3), 579–592. https://europepmc.org/article/MED/21744974

Sutin, A. R., et al. (2016). The five-factor model of personality and physical inactivity: A meta-analysis of 16 samples. *Journal of Research in Personality, 63*, 22–28. https://doi.org/10.1016/j.jrp.2016.05.001

Swider, B. W., & Zimmerman, R. D. (2010). Born to burnout: A meta-analytic path model of personality, job burnout, and work outcomes. *Journal of Vocational Behavior, 76*(3), 487–506. https://doi.org/10.1016/j.jvb.2010.01.003

Tam, K.-P. (2013). Concepts and measures related to connection to nature: Similarities and differences. *Journal of Environmental Psychology, 34*, 64–78. https://doi.org/10.1016/j.jenvp.2013.01.004

Tatar, A. (2017). [Turkish translation of the Big Five-50 Personality Test and its comparison with the short form of the Five Factor Personality Inventory]. *Anadolu Psikiyatri Dergisi, 18*(1), 51–61. https://search.trdizin.gov.tr/yayin/detay/266100 (reference to be verified)

Tatar, A. (2018). [Psychometric properties of the Turkish long and short forms of the HEXACO-PI-R]. *Anadolu Psikiyatri Dergisi, 19*(Suppl. 1). https://www.imrpress.com/journal/AP/19/Supplement%201/10.5455/apd.283098 (reference to be verified)

Teixeira, G. P., Mota, M. C., & Crispim, C. A. (2018). Eveningness is associated with skipping breakfast and poor nutritional intake in Brazilian undergraduate students. *Chronobiology International, 35*(3), 358–367. https://europepmc.org/search?query=%22Eveningness%20is%20associated%20with%20skipping%20breakfast%22

Terracciano, A., & Costa, P. T. (2004). Smoking and the Five-Factor Model of personality. *Addiction, 99*(4), 472–481. https://pmc.ncbi.nlm.nih.gov/articles/PMC2376761/

Tett, R. P., Fitzke, J. R., Wadlington, P. L., Davies, S. A., Anderson, M. G., & Foster, J. (2009). The use of personality test norms in work settings: Effects of sample size and relevance. *Journal of Occupational and Organizational Psychology, 82*(3), 639–659. https://doi.org/10.1348/096317908X336159

Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy, 112*(S1), S164–S187. https://doi.org/10.1086/380085

The Myers-Briggs Company. (n.d.). *Trademarks*. https://themyersbriggs.com/en-US/Support/Trademarks

Thomas, D. M., et al. (2014). Effect of dietary adherence on the body weight plateau: A mathematical model incorporating intermittent compliance with energy intake prescription. *American Journal of Clinical Nutrition, 100*(3), 787–795. https://doi.org/10.3945/ajcn.113.079822

Tracey, T. J. G. (1997). RANDALL: A Microsoft FORTRAN program for a randomization test of hypothesized order relations. *Educational and Psychological Measurement, 57*(1), 164–168. https://doi.org/10.1177/0013164497057001012

Traitify. (n.d.). *The science*. https://www.traitify.com/science

Truity. (n.d.). *TypeFinder personality test*. https://www.truity.com/test/type-finder-personality-test-new

Trustpilot. (n.d.). *16Personalities reviews*. https://ca.trustpilot.com/review/16personalities.com

Tsaousis, I. (2010). Circadian preferences and personality traits: A meta-analysis. *European Journal of Personality, 24*(4), 356–373. https://doi.org/10.1002/per.754

Turiano, N. A., et al. (2012). Personality and substance use in midlife: Conscientiousness as a moderator and the effects of trait change. *Journal of Research in Personality, 46*(3), 295–305. https://pmc.ncbi.nlm.nih.gov/articles/PMC3388488

Twomey, C., & Johnson, J. A. (2022). Open-source personality trait norms for the United Kingdom and Ireland. *European Journal of Psychological Assessment, 38*(6), 432–439. https://doi.org/10.1027/1015-5759/a000644

University of Connecticut. (n.d.). [Adding prescriptive feedback to written personality feedback; doctoral dissertation, n = 132]. https://digitalcommons.lib.uconn.edu/dissertations/AAI3504792 (reference to be verified)

University of Oxford. (n.d.). False personality feedback and mood [Thesis]. https://ora.ox.ac.uk/objects/uuid:c88e03b3-19ef-4929-9b06-add78422de06 (reference to be verified)

University of Warwick. (n.d.). The Barnum effect and behavioural tendencies [Doctoral thesis]. https://wrap.warwick.ac.uk/id/eprint/96198 (reference to be verified)

Uygur, H., Tekdemir, R., Uygur, Ö. F., Aydın, E. F., Çelik, M., Babacan, H. E., et al. (2024). Psychometric properties of the Turkish reduced Morningness-Eveningness Questionnaire. *Chronobiology International, 41*(5), 632–646. https://doi.org/10.1080/07420528.2024.2339964

Vardarlı, B., Özyürek, R., Wilkins-Yel, K. G., & Tracey, T. J. G. (2017). Examining the structure of vocational interests in Turkey in the context of the personal globe model. *International Journal for Educational and Vocational Guidance, 17*(3), 347–359. https://doi.org/10.1007/s10775-016-9338-6

Vedel, A., Wellnitz, K. B., Ludeke, S., Soto, C. J., John, O. P., & Andersen, S. C. (2021). Development and validation of the Danish Big Five Inventory-2: Domain- and facet-level structure, construct validity, and reliability. *European Journal of Psychological Assessment, 37*(1), 42–51. https://doi.org/10.1027/1015-5759/a000570

VIA Institute on Character. (n.d.). *Terms of service*. https://www.viacharacter.org/terms-of-service

Villar, A., Callegaro, M., & Yang, Y. (2013). Where am I? A meta-analysis of experiments on the effects of progress indicators for web surveys. *Social Science Computer Review, 31*(6), 744–762. https://openaccess.city.ac.uk/id/eprint/14427/

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology, 74*, 577–596. https://papers.ssrn.com/abstract=4337272

Wasti, S. A., Lee, K., Ashton, M. C., & Somer, O. (2008). Six Turkish personality factors and the HEXACO model of personality structure. *Journal of Cross-Cultural Psychology, 39*(6), 665–684. https://doi.org/10.1177/0022022108323783

Webb, T. L., Christian, J., & Armitage, C. J. (2007). Helping students turn up for class: Implementation intentions and conscientiousness. *Learning and Individual Differences, 17*(4), 316–327. https://eprints.whiterose.ac.uk/id/eprint/10755

Welz, M., & Alfons, A. (2023). *When respondents don't care anymore: Identifying the onset of careless responding* [Preprint]. arXiv. https://arxiv.org/abs/2303.07167

Westenhoefer, J., Stunkard, A. J., & Pudel, V. (1999). Validation of the flexible and rigid control dimensions of dietary restraint. *International Journal of Eating Disorders, 26*, 53–64. https://europepmc.org/article/MED/10349584

White, B. X., Jiang, D., & Albarracín, D. (2021). The limits of defaults: The influence of decision time on default effects. *Social Cognition, 39*(5), 543–569. https://doi.org/10.1521/soco.2021.39.5.543

White, M. P., et al. (2019). Spending at least 120 minutes a week in nature is associated with good health and wellbeing. *Scientific Reports, 9*, 7730. https://doi.org/10.1038/s41598-019-44097-3

Whiteside, S. P., & Lynam, D. R. (2001). The Five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*(4), 669–689. https://doi.org/10.1016/S0191-8869(00)00064-7

Wilmot, M. P., & Ones, D. S. (2019). A century of research on conscientiousness at work. *Proceedings of the National Academy of Sciences, 116*(46), 23004–23010. https://doi.org/10.1073/pnas.1909972116

Wilson, K. E., & Dishman, R. K. (2015). Personality and physical activity: A systematic review and meta-analysis. *Personality and Individual Differences, 72*, 230–242. https://doi.org/10.1016/j.paid.2014.08.023

Xu, Y., Beller, A. H., Roberts, B. W., & Brown, J. R. (2015). Personality and young adult financial distress. *Journal of Economic Psychology, 51*, 90–100. https://experts.illinois.edu/en/publications/personality-and-young-adult-financial-distress/

Zelenak, M. F. (2016). *2014 American Community Survey (ACS) Internet test* (American Community Survey Research and Evaluation Report Memorandum Series No. ACS16-RER-03). U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2016/acs/2016_Zelenak_01.pdf

Zelenski, J. M., Santoro, M. S., & Whelan, D. C. (2012). Would introverts be better off if they acted more like extraverts? *Emotion, 12*(2), 290–303. https://doi.org/10.1037/a0025169

Zhao, H., & Seibert, S. E. (2006). The Big Five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology, 91*(2), 259–271. https://doi.org/10.1037/0021-9010.91.2.259

Zhao, H., Seibert, S. E., & Lumpkin, G. T. (2010). The relationship of personality to entrepreneurial intentions and performance: A meta-analytic review. *Journal of Management, 36*(2), 381–404. https://doi.org/10.1177/0149206309335187

Zheng, Y., et al. (2015). Self-weighing in weight management: A systematic literature review. *Obesity, 23*(2), 256–265. https://doi.org/10.1002/oby.20946

Appendices

Appendix A. Index of decision records (ADR)

The decision records cited in this monograph are under `docs/decisions/`. Each record carries the sections Question, Decision, Rationale, Rejected alternative and Impact; `Back-end/tests/Feature/InvariantsTest.php` makes some of the decisions executable.

Record	Title	Chapter
ADR-003	Application Time Zone	8
ADR-004	Streak Semantics	6, 9
ADR-010	Activity Log, Derived Goals and Goal Completion	6
ADR-013	Nutrition Plan and Food Memory	6
ADR-016	A Single Source for Derived Values — the Link between Widgets	5
ADR-017	Pill Box — Medication and Supplement Plan	8
ADR-019	Notifications — Pill Reminders and the Notification Panel	6
ADR-024	Budget — Three Envelopes, "Safe to Spend" and the Burden of Logging	6, 8
ADR-025	Pusula Personality Test — Pre-Sign-Up Value and a Separate Service	1, 2, 5, 6, 7, 8
ADR-026	Personality Test Form v4 — Two Stages, Derived Types, E-mail Close to the Result	1, 2, 3, 4, 5, 8
ADR-027	Pusula-API — Pusula's Core Forked into Rehnüma	1, 8
ADR-028	Report Depth — Mind Motif, Temperament and Expandable Personal Sections	1, 2, 3, 4, 5, 7, 8, 9
ADR-029	Personality Test Validation Gates and Simulation Findings	1, 2, 3, 4, 5, 6, 7, 8, 9
ADR-030	Item Keying Balance and Acquiescence Bias	2, 3, 4, 8
ADR-031	Character Analysis — Design Language and Trust Framing	3, 5, 6, 7, 8
ADR-032	Paged Report, Reference-Group Percentiles and Universal Codes	3, 4, 5, 6, 7
ADR-033	Character Trace — Evidence Map and Rule Engine	1, 2, 5, 6, 7, 8, 9
ADR-034	Five Pillars — the Life Tree Widget	7
ADR-035	Goal Widget Family, Sprint and Progress Trace	6, 7
ADR-037	Five-Branch Character Matrix — the Rehnüma Standard	5, 7, 10
ADR-038	Calendar and Habits on the Tasks Page	6
ADR-041	Sales Model and Launch Price	1, 5, 8
ADR-044	Form v5 — Self-Placement, Statement Ranking, Visible Consistency	1, 3, 4, 5, 6, 9, 10
ADR-045	Report Guide — New Sections and Writing Rhythm	5, 7
ADR-046	Five Languages — Server-Side Localisation	1, 3, 5, 6, 7, 8, 9
ADR-055	Default language English, routes in English	8

Record	Title	Chapter
ADR-058 (`rapor-dili-is-tek-dilinde`)	Analysis and report in the request language; language selector visible	5
ADR-058 (`yasal-sayfalar`)	Legal Pages State What the Application Actually Does; Sources Page	8
ADR-061	Pricing Page: Three Cards, Dollars; Report One-off, Dashboard Annual	5
ADR-062	On-site Purchase; Payment through PayTR, the Order Is Paid on Notification	8
ADR-063	Campaign Price; the Admin Panel Is the Single Source of the Price; Report and Dashboard as Two Products	8
ADR-064	Transactional E-mails; Password Reset; New-Device Alert	8
ADR-065	Sign-in with Google	5
ADR-067	Two-Step Verification (E-mail Code), Verification Reminder, Account Deletion	8
ADR-073	The Analysis Result Is Unlocked with an Account and Shown on the Dashboard	1, 2, 3, 5, 6, 8
ADR-074	The Target Occupation Is Chosen at the Start of the Test and Its Fit Is Shown in the Report	3, 5, 8
ADR-080	A Single Design System for the Analysis Side: Top Bar, Label, Minimum Type Size, Warm Dark	5
ADR-081	Report in Book Layout; the Locked Section Opens, the Remainder Is Blurred	2, 5, 7, 8, 9
ADR-082	Personality Percentiles from the Published Norm (IPIP-NEO-120, Johnson 2014)	2, 3, 4, 5, 6, 7, 8, 9, 10
ADR-083	Price in Lira, Two Tiers: Free Analysis and Annual Dashboard	1, 5, 8, 9

Appendix B. Research notes and reports

Location	Content
`docs/research/reports/Rehnüma kişilik testi psikometri.md`	Instrument selection, scoring and report language, competitor products (6 notes and one direct-observation record)
`docs/research/reports/Kişilik testi doğrulama protokolü.md`	Validation gates, translation review, derived layers (3 notes)
`docs/research/reports/Karakter izi kanıt haritası.md`	Six-strand evidence review, evidence policy, rule schema (6 notes)
`docs/research/reports/Beş dal karakter matrisi.md`	Review of personality and behaviour in five life branches, the basis of the corpus (5 notes)
`docs/research/research_notes/`	21 files in the four folders on which this monograph rests; every finding is sourced
`docs/research/pilot/`	Cognitive interview protocol, participant texts, reconciliation table

Appendix C. Reproduction

Simulation (Gate 0). With Pusula-API running locally (`php artisan serve --port=8100`, with `PUSULA_RATE_PER_MINUTE=100000` in `.env`):

```
cd docs/research/scripts
export PUSULA_SERVICE_KEY=…
python simulate.py --n 300 --retest 60 --out ./sim
```

Note. The Gate 0 run was made with form v4. The service today serves form v5 (`Pusula-API/app/Assessment/FormBuilder.php`); `simulate.py` answers only the section types of v4 and does not recognise the `self_place` and `rank` steps of v5. To repeat the run, the script has to be adapted to these two steps.

Psychometric analysis (Gates 1–4).

```
cd Pusula-API && php artisan pusula:export-responses storage/app/export.jsonl --completed
python ../docs/research/scripts/psychometrics.py ../Pusula-API/storage/app/export.jsonl --
stage 2 --json rapor.json
```

Validation of the rule catalogue and the five-branch corpus.

```
cd Back-end && php artisan test --compact --
filter='PersonalityRule|DomainMatrix|Invariants'
```

Compiling the monograph.

```
cd docs/research/tez/tooling && npm install && node build.mjs
node build.mjs --lang en   # English edition
```

Appendix D. Sample rule catalogue record

A record conforming to the contract in `Back-end/database/data/personality_rules/README.md` (`weightTrend.json`):

```
{
  "id": "weightTrend.S5.plateau-neutral",
  "widget": "weightTrend",
  "scenario": "S5",
  "decision": "K13",
  "when": { "all": [ { "var": "domain.S", "band": "low" } ] },
  "effect": { "type": "copy", "key": "weight.plateau.note", "variant": "gentle", "value":
null },
  "copy": { "en": "The trend has been flat for a while. This is seen in most people losing
weight around month 6 and is usually the sum of small daily deviations; that is how the
body works. The next step: fix one meal time, not your goal.", "why": "Plato beklenen bir
dönem; suçlamasız dil düzelme motivasyonunu destekliyor (BreinesChen2012)." },
  "evidence": { "grade": "orta", "sources": ["Thomas2014", "BreinesChen2012",
"Elfhag2008"], "note": "Plato mekanizması ve öz şefkat çerçevesi kişilikten bağımsız
kanıt; düşük S'de ek fayda çıkarım. Yalnızca yumuşatma; bilgi gizlenmez, 'uyumsuzluk'
sözcüğü kullanılmaz. (Rapor §6 örnek 1.)" },
  "status": "rule"
}
```


The `copy.en` string is the product's English text for this rule (`Back-end/database/data/en/personality_rules/copy.json`); the remaining Turkish values are kept as they stand in the catalogue: `why` reads "A plateau is an expected period; blame-free language supports the motivation to recover (BreinesChen2012)", the evidence `grade` "orta" is "moderate", and the evidence `note` reads "The plateau mechanism and the self-compassion framing are personality-independent evidence; the additional benefit at low S is an inference. Softening only; no information is hidden, and the word 'non-adherence' is not used. (Report §6, example 1.)"

Appendix E. Public documents

Page	Path	Content
Why you can trust it	`/analysis/trust`	Instruments, validation gates and their status, honesty principles, limits, data; a link to this monograph
Instruments and changes	`/analysis/changes`	The original and the Turkish of the 120 IPIP items, 60 interest items and 20 value cards, with the change log; the O*NET §3(b) notice
Sample report	`/analysis/report/sample`	The full report of a synthetic profile, labelled "not a real person"

Note. The routes are in English and the default language, English, is served without a prefix; Turkish pages open with the `/tr` prefix (e.g. `/tr/analysis/trust`; ADR-055).