


UNITED WISDOM

# Personality tests: what a result can and cannot tell you

The Person and Inner Life

Andrey Zarinsh & Maria Evilkut

I-04-N04 · EN

@andrey.zarinsh @ilkutki

# Personality tests: what a result can and cannot tell you

## Epigraph

> *“Know thyself!”*

Plato, *Charmides*, 164d: the Delphic inscription in Critias's argument, in Benjamin Jowett's translation. The subsequent dialogue questions the account of self-knowledge proposed here.

## What this article is about

Imagine a hypothetical situation: someone completes a questionnaire and recognises themselves in the description. It becomes easier to explain their love of quiet, habit of preparing in advance, or tiredness after social gatherings. Then the result starts leading a life of its own: “That kind of conversation doesn't suit me,” “I can't lead,” “This person is a different type; we'll never agree.” A handful of answers becomes permission, or a prohibition, governing an entire life.

A good instrument can help frame a question about habits. Yet the steps between an answer, a number, a trait description, and a life decision need justification. This article considers them separately: what a scale measures, who supplies the comparison, where confidence in an interpretation comes from, and how to examine it through action. Questionnaire research is placed alongside a philosophical discipline of self-knowledge; that comparison does not make an ancient text evidence for a psychometric model. Our aim is to preserve measurement's usefulness and a person's freedom to understand themselves more carefully, without substituting one for the other.

## I. A result begins with a particular question

The name “personality test” creates an expectation that an instrument examines a whole person. But we always have a particular instrument in front of us. One gathers self-reports about typical habits, another assesses performance on a task, and a third serves a specialised professional assessment. These procedures cannot be grouped together simply because a number appears at the end. Before reading the description, it helps to establish the name and version, purpose, language, instructions, and grounds for interpretation. Uncertainty about those conditions makes an inference less definite, even when the presentation looks convincing.



@andrey.zarinsh @ilkutki

In a hypothetical example, a question about enjoying discussions of ideas in a group concerns the person's response to a particular wording. It does not directly ask whether they can support someone close to them, speak to an audience, or deserve trust. Moving between these subjects requires a separate basis. A dedicated study might find a relationship, but merely extending the meaning of “sociability” cannot establish it. That step is especially easy to take when the description uses appealing, almost biographical language and offers a ready-made portrait of its central character.

First, we can state the fact closest to the procedure: “I selected these answers under these instructions.” Then we can discuss the score according to that instrument's rules. Only after that does the question arise of how far it relates to everyday actions. This sequence does not devalue the test. It reveals where processing answers ends and interpretation begins. Someone is entitled to recognise themselves in a description, but recognition alone does not establish a method's accuracy: different texts can produce a similar impression.

If a hypothetical result describes a preference for order, the first useful question is concrete: in which activities does it appear, and whom does it help? A question about the person's “true essence” immediately asks too much. Sometimes the result identifies a familiar habit; sometimes it draws attention to a distinction previously overlooked; sometimes it remains too general. In all three cases, the instrument can be judged by its actual contribution rather than by a promise to reveal a person conclusively to themselves.

## II. A trait is not a box containing people

In the documentation for IPIP, an open collection of personality measures, the authors specifically warn against turning scores into rigid categories. A scale may describe a position on a continuous dimension. A boundary between “low” and “high” then serves as a way to report a result, rather than necessarily marking two naturally distinct kinds of people. The official IPIP explanation should be read alongside documentation for the selected scale: the collection does not provide a single manual that could replace research on each instrument.

Imagine hypothetical values on either side of an established boundary. If one point changes the category's name, we cannot automatically conclude that the person's entire way of being has changed. Two neighbouring results can be closer to each other than two values within the same broad category. At the same time, categories are convenient in conversation: a short word is easier to remember. Convenience becomes a problem when we attribute more differences to that word than the underlying scale contains and forget that its division is conventional.

There is also a distinction between a tendency and inevitability. A person may more often choose familiar surroundings and still take an enthusiastic part in a particular trip. An exception

does not destroy every possibility of speaking about habits. But a habit does not prohibit an exception either. Discussing a trait requires both levels: a sufficiently consistent pattern and the conditions under which it appears differently. A generalisation gains substance when it allows clarification, rather than defending itself against every example that fails to fit.

Careful language here is therefore more than politeness. “I more often recover my energy in quiet surroundings” leaves room to examine frequency and conditions. “I'm that type of person” can easily become a closed explanation: every action is interpreted in advance as confirmation. Then the possibility of learning disappears. A label can be a useful abbreviation if someone remembers what it abbreviates. When it starts determining abilities, relationships, and worth, we need to return to observations and examine each extension separately.

## III. Who supplies the comparison for a number?

A number looks like an independent fact, but its meaning often depends on a comparison. We need to distinguish a sum of answers, an average, and a position among members of a particular group. A percentile refers to a place in a chosen distribution, rather than the proportion of a personality occupied by a quality. A hypothetical “eighty” does not mean that someone is eighty percent kind. Without an explanation of what has been calculated, an intelligible figure can support an entirely unintelligible conclusion that then seems mathematically confirmed.

IPIP's recommendation to compare scores with an appropriate sample underscores the need to know where norms come from. The comparison group may differ in age, language, participation conditions, and other relevant respects. That does not make every transfer impossible, but it calls for justification. Comparison with visitors to one website and comparison with a country's population answer different questions. If the group's composition is not given, we cannot silently choose the most impressive interpretation and speak as though the result itself had established it.

Imagine a hypothetical person scoring above average on a scale within a small study group. That result does not reveal their position among all adults, or explain why they answered as they did. The comparison may be useful for a local task; it may be too narrow for the intended decision. The broader the decision, the more important it is to check whether its basis fits. A participant can ask what “above” means without treating that question as distrust of their own experience or of the questionnaire's creators.

Comparison also has an ethical dimension. A position in a distribution is not a place in a queue for respect. A high score is not always better than a low one: its significance depends on the quality, task, and consequences under consideration. Even where an ability is useful, assessing one ability does not rank whole human beings. In practice, a number needs to retain

its unit, comparison group, and the question for which it was obtained. If those details are missing, pausing interpretation becomes a substantive part of understanding, rather than an obstacle to it.

## IV. Reliability and meaning are different questions

To assess an instrument, it is not enough to ask whether it produced a similar result twice. Repeatability matters, but a repeatable procedure can consistently record something other than what a particular decision requires. A simple distinction helps: how consistently scores are obtained, and how well their interpretation is justified for a chosen purpose. These are different questions about measurement quality. Answering the first does not settle the second; an attractive interpretation does not replace checking the first either. The terms serve clarity, rather than creating an impression of professional authority.

In Soto and John's BFI-2 study, developing the inventory included examining its structure and its relationships with other measures. The authors describe five broad domains and fifteen narrower facets. This is a particular research programme, rather than a promise that every internet test with a similar name underwent the same checks. A reference to the “Big Five” alone does not transfer BFI-2's properties to another translation, set of questions, or scoring method. The version actually used needs its own documentation.

Next, we need to ask which interpretation was examined. An association with a similar scale, an account of response structure, and a prediction of a particular behaviour are not one and the same piece of evidence. If a hypothetical report proposes choosing a career, there must be grounds for that use specifically. A general statement that personality is studied scientifically cannot fill the gap. The scientific standing of a field does not make every practical promise within it established. Nor does a lack of relevant evidence establish the opposite: it means that the proposed conclusion remains broader than its available basis.

This discipline avoids arguing simply about faith in tests as a whole. We can recognise a particular questionnaire's usefulness for a particular description and reject its use for another task. A hypothetical person may gain material for discussing habits without gaining grounds to predict a marriage's success. The more precisely the purpose is stated, the easier it is to assess the instrument's quality. One enormous question, “Is the test true?”, becomes several examinable ones: what it collects, which conclusions it supports, under what conditions, and with what uncertainty.

## V. An answer about yourself has its own conditions

Self-report gives access to how someone understands and describes themselves. It is an important source because some inner experience is not directly visible to other people. At the same time, an answer arises in a specific situation: someone reads particular words, recalls selected episodes, and decides what counts as typical. These conditions do not make them an unreliable witness by default. They simply mean that a statement about oneself cannot be treated as an observation of a whole life without further clarification. Sincerity and completeness are different properties of an account.

In a hypothetical example, one participant thinks of work meetings, another of gatherings with people close to them. Both answer a question about enjoying social contact, but they are drawing on different relationships. Perhaps the instructions specify a context; perhaps they leave it broad. If the result comes as a surprise, returning to that distinction can help. The issue may be which part of experience was used to answer, rather than a hidden quality. Examination starts with the meaning of the chosen examples, rather than a suspicion that the result has exposed a concealed truth.

Here it is especially important to distinguish feeling from outward action. Someone may report that public speaking requires considerable effort and still speak well. Others may hear calm speech without knowing its inner cost. The two accounts can complement each other. Neither has to cancel the other if they concern different questions. But experiencing effort does not establish inability, and giving a successful presentation does not establish an absence of tension. A good interpretation preserves the distinction rather than forcing a choice between the accounts.

If a result produces shame or pride, we can notice that experience without declaring it evidence of accuracy. A strong reaction shows the description's significance for the person; it does not validate the procedure. Sometimes a particularly relevant question is what permission the person wanted to receive: to decline an unpleasant activity, explain a conflict, or acknowledge a need for rest. Those needs deserve direct discussion. A score does not have to become the intermediary that alone decides whether someone is entitled to respect, support, or a change in their familiar surroundings.

## VI. A match with life needs clarification

After reading a result, finding a confirming episode is easy. In a hypothetical situation, a participant recognises a description of a preference for planning and remembers preparing an itinerary in advance. The connection makes sense, but one match is insufficient for a broad conclusion. We need to ask how often it occurs, under which circumstances, and which other explanations fit. Perhaps the journey required preparation because of complicated connections; perhaps the person prefers plans across many activities. These accounts do not automatically exclude each other and require different material to examine.

It helps to look for sufficiently specific cases rather than putting the person on trial. If they want to organise work better, they can discuss recent, comparable tasks: what helped them start, what prevented completion, and how much uncertainty was acceptable. A description thereby becomes a working hypothesis. It gains the possibility of changing with new information. “A short plan helps me before a complex task” is much easier to examine than a grand assertion that one belongs to a special kind of person, which explains everything and therefore distinguishes almost nothing.

Cases that do not match also deserve to be retained. They may reveal the limit of a generalisation, a different context, or the particular importance of a goal. If a hypothetical lover of order tolerates disorder while creating something, that can become a meaningful question about priorities. We need not immediately declare the episode an exception that is forbidden to count. Nor does one episode require abandoning the whole pattern. A good examination retains proportion: a limited observation refines a limited hypothesis. It becomes neither a final exposure of error nor a final victory of a description over life.

A result is especially useful when it reveals a distinction previously merged into one problem. A hypothetical person might discover that the difficulty is not social contact in general but a conversation without a clear purpose. That clarification can support a change in meeting format. Yet it emerges through considering circumstances together, rather than being automatically deduced from a number. The test participates in a search for language. The next step belongs to the person: compare the wording with experience, discuss consequences, and choose an action that reflects actual conditions rather than merely an expressive report heading.

## VII. Charmides: knowing a label is not knowing yourself

The Delphic invitation to self-knowledge often sounds like a promise to find the right definition of a person. In Plato's *Charmides*, however, the discussion is more complicated. Critias connects self-knowledge with temperance, while Socrates tests proposed definitions through questions. Can someone know that they know without knowing the subject? How would such knowledge distinguish a competent physician from someone who merely appears competent? The dialogue requires clarification of knowledge's content and usefulness. A formula does not remove the need for further inquiry.

This is philosophical reasoning about knowledge and virtue, rather than an ancient description of modern scales. We use it for a limited comparison: possessing a label verbally does not establish which subject has actually been understood. If a hypothetical reader can name their profile, the question remains of what they have learned about actions, relationships, and possibilities. Perhaps they have gained a useful distinction. Perhaps they have acquired only a new word for an old worry. Content determines the answer, rather than the antiquity of a slogan or the modernity of the interface displaying a result.

It matters that Socrates is not credited with simply endorsing Critias's thesis. The text conducts an argument rather than dispensing a ready-made procedure for defining oneself. Its tension deserves to remain: a wish for clarity encounters the difficulty of establishing one's own knowledge. In life, a similar tension can be accepted without rejecting instruments' assistance. Someone can use a description while acknowledging that they do not yet understand all its implications. An acknowledged boundary of knowledge sometimes enables more precise action than an overconfident explanation that closes questions before they have been considered.

The practical conclusion here belongs to this article's authors: after a label, return to the subject. What specifically has become clearer? Which observation supports that understanding? Which task does it help address? If answers are missing, the description can remain a subject for reflection without being presented as completed self-knowledge. This approach does not require endless doubt about every choice. It requires proportion between an assertion's strength and its basis. A person remains more than any description because life continues to bring new circumstances and questions.

## VIII. A description does not deliver a moral verdict

A scale may describe a habit, but respect for a person should not depend on their profile's attractiveness. In a hypothetical report, a term associated with emotional sensitivity may sound reproachful, while another sounds like praise. Yet a score's meaning is established by documentation rather than by the everyday moral colouring of its name. It is important not to turn description into accusation or exoneration. We cannot conclude that someone is bad because they received a particular score, or that they may cause harm because a description recognised their tendency towards sharpness.

Here, explaining conditions differs from responsibility for action. If a hypothetical participant struggles with long discussions, people can agree on a break. If they humiliate someone else, the action and its consequences need discussion. A test can suggest a subject for clarification but cannot replace agreements about acceptable treatment. Conversely, an accusation concerning a general quality does not clarify the specific change required. Proportionate feedback names an action, a condition, and an expectation, leaving room for a substantive response.

The same boundary matters in discussions of compatibility. Two descriptions cannot establish whether people can listen to one another, agree about money, or respect a refusal. Those questions emerge in particular relationships. Differences in habits may ease a discussion of everyday arrangements: one person needs a plan, another room for change. Yet the existence of a difference predicts neither failure nor success. A conclusion requires seeing how the participants handle it. A short profile cannot replace mutual agreement or observation of actual arrangements.

This article does not offer clinical interpretation of scores. If an instrument claims a diagnostic purpose, its professional requirements and an appropriate specialist assessment apply; a popular description cannot acquire that function by association. It is particularly important not to explain serious difficulties through one label or use this article for self-diagnosis. A person can acknowledge that a result touched a painful subject and seek suitable support without establishing a problem's origin from a few answers. Care begins with the person's situation rather than the convenience of classification.

## IX. A life decision needs a broader basis

A result can contribute to a decision, but the weight of that contribution depends on consequences. Choosing a convenient note-taking format calls for only a small, reversible experiment. Decisions about education, employment, or relationships require broader information. If a hypothetical person declines to apply for an interesting position solely because of a short profile, they close a possibility that has not yet been examined. Perhaps the job really is unsuitable. But responsibilities, conditions, skills, interests, and experience help establish that, rather than one trait label.

It matters to distinguish a group prediction from an individual's destiny. Even a supported relationship between a score and an outcome does not describe every circumstance of a particular life. A specific decision requires additional information and consideration of uncertainty. A scientific expression of a probabilistic relationship cannot become an unconditional “you will not succeed.” If a report makes a strong promise, it helps to request grounds for that promise specifically. The more serious the consequences, the less suitable an explanation resting only on measurement's general authority and an appealing language of certainty becomes.

There is also a question of consent: who obtained the result, who sees it, and why is it being used? Discussing a voluntary self-description differs from another person's decision about hiring, admission, or a relationship. Before sharing results, it helps to know which information is necessary and how it will be used. Not every personal answer needs publication to demonstrate good faith. This article does not establish legal rules for data processing. Its practical point is simpler: personal information deserves a definite purpose, intelligible handling, and an opportunity to ask questions before an inference affects someone's life.

Proportion makes it possible to use a test without either extreme. We do not have to reject every measurement to protect freedom. We do not have to give measurement the final word to acknowledge its value. A hypothetical participant can try a suitable working method, compare educational options, and discuss needs with people close to them. In each case, they clarify what the result supports and which decisions require other material. That position preserves responsibility: an instrument supplies information within a procedure's limits, while the person considers its significance among the other grounds for their choice.

## X. A working description must allow change

A good description has a date, context, and possibility of revision. This does not mean that every score must change quickly. It means that one result does not describe a person for all future time. If a hypothetical reader returns to a questionnaire later, the version, instructions, and comparison conditions matter. A changed score alone does not explain its cause. Such an explanation requires additional consideration; we cannot automatically declare it growth, deterioration, healing, or a revelation of authentic character.

In everyday life, observing an action that actually matters to the person is more useful. If they wanted to learn to express their needs, they can consider whether doing so became easier in a particular conversation. If they wanted to organise a working day, they can assess execution of a realistic plan. That information need not fit an attractive story about a changing profile. It retains its own value. A test can be one reference point, but a chosen action's success need not wait for a number to permit the person to call it successful.

In a hypothetical situation, a description helped a reader stop demanding constant sociability from themselves. That can be useful if they understand their need to recover more accurately and respect their relationships. But if the description prohibits every gathering requiring effort, its usefulness becomes questionable. The boundary concerns whether the language expands understanding of conditions or closes access to new experience. One check of a category's correctness cannot settle that: we must consider the consequences of application, which arise in real life and affect other participants.

An eventual self-description can therefore be modest and substantive: “On this scale, with these answers, I received this score. It fits several habits but does not explain everything. I will try taking one of them into account in a particular task.” That statement contains information, a boundary, and a next step. Self-knowledge does not end with a report. It continues through questions, actions, feedback, and an ability to revise explanations. An instrument helps when it supports that process and a person retains the right to be more attentive and complex than their result.

## What this means in practice

Choose a result already obtained voluntarily that can be discussed without pressure. Record the instrument's name, version, language, and purpose. Find an explanation of the chosen scale and establish whether the number means a sum of answers, an average, or a comparison with a group. If documentation is absent, limit the conclusion to that fact: the score's precise meaning has not yet been established. Do not replace a missing explanation with the most convincing text from another source.

Take one assertion from the report. Turn it into a question about a particular action: when does it occur, what helps, and where are the exceptions? Find a few ordinary, sufficiently safe episodes. Do not seek confirmation at any cost or organise a search for hidden causes. You may discover a match, a qualification, or wording that is too broad; each helps identify the description's boundary.

If you want to change something, choose a small, reversible step. In a hypothetical example, someone who likes preparation could request a meeting agenda beforehand and see whether it helps them participate. Assess the action through its consequences for the task and relationships. Then decide whether to continue. For serious life decisions, add information about actual conditions, experience, and available alternatives. A questionnaire can offer language for a conversation; entitlement to respect and support does not require a suitable profile.

## Key quotation

> *“Be temperate!”*

Plato, *Charmides*, 164e: a formula in Critias's interpretation of the Delphic greeting, in Benjamin Jowett's translation. The speaker matters: Socrates subsequently investigates whether temperance should be understood as self-knowledge. The quotation is not the dialogue's final definition and does not validate modern tests.

## Connections between traditions

| Approach | What is compared | Similarity | Limit of the comparison |
|---|---|---|---|
| IPIP | Interpreting a position on a scale | A description's meaning needs clarification | A psychometric scale does not deliver a moral verdict |
| BFI-2 | Broad domains and narrower facets | A general label requires substantive distinctions | A research structure is not a complete personality portrait |
| Plato, Charmides | Testing a definition of self-knowledge | Naming knowledge does not replace knowing its subject | A philosophical argument about temperance does not validate a questionnaire |
| The Delphic address in Critias's speech | An invitation to attend to oneself | Self-understanding requires the person's participation | An ancient religious and ethical address is not a measurement procedure |
| The authors' practical position | Clarification through actions and consequences | A description remains open to revision | Everyday examination does not replace scientific validation or clinical assessment |

## Primary sources and research

• IPIP. Interpreting Individual IPIP Scale Scores. Official explanation of norms, continuous scales, and the limits of categories; the complete page was read. Any specific instrument requires its own research and documentation.

• Christopher J. Soto, Oliver P. John. The Next Big Five Inventory (BFI-2), 2017. Journal of Personality and Social Psychology, 113(1), 117–143. The abstract, introductory account of aims, and General Discussion were read. Used for the instrument's design and research boundaries, without reproducing questionnaire items or independently analysing all underlying data.

• Plato. Charmides, Benjamin Jowett translation. Project Gutenberg electronic edition. The discussion of self-knowledge and knowledge of knowledge, 164d–172c, was considered; Critias's statements are distinguished from Socrates's questions. The short Russian quotations are the authors' translations from this English version.

## Tags

self-knowledge, personality test, questionnaire, self-report, psychometrics, IPIP, BFI-2, Big Five, personality traits, scale, norms, percentile, interpretation, reliability, validity, context, feedback, habits, life choices, freedom, responsibility, comparison, categories, Plato, Charmides, Critias, Socrates, temperance, limits of knowledge, United Wisdom

## About the authors

Andrey Zarinsh is the author and founder of United Wisdom. He explores inner transformation, the unity of traditions, intercultural dialogue, and contemporary spiritual discipline. Instagram: @andrey.zarinsh.

Maria Evilkut is a coauthor of United Wisdom. Alongside her research, she seeks to test these ideas in practice and observe how principles of unity and mutual understanding work in real life between people and cultures. Instagram: @ilkutki.