*O. V. Shcherbakova, D. N. Makarova, E. A. Nikiforova*

# "CONCEPT SYNTHESIS" TEST FOR INTELLIGENCE RESEARCH: QUALITATIVE ANALYSIS AND CONCEPTUAL COMMENTARIES

"Concept Synthesis" test (by M. A. Kholodnaya) was suggested within an ontological approach to intelligence studies as an alternative tool for measuring intellectual abilities. The aim of this paper is to summarize our experience of using this test as well as to provide some conceptual commentaries on the original criteria developed for rating the answers. We compiled a database of 924 answers for "Concept Synthesis" test, collected from 2013 to 2015 in three different studies (all together, 104 participants) and rated in accordance with the original criteria. All the answers underwent a thorough qualitative analysis and were re-rated independently by three experts. The updated criteria for assessing the answers were suggested as a result of this work. We also provide conceptual commentaries on more than 40 particular answers and discuss the most common technical issues emerging while using "Concept Synthesis" test. Refs 19. Table 1.

*Keywords:* intelligence, intelligence measurement, conceptual abilities, "Concept Synthesis" test.

*О. В. Щербакова, Д. Н. Макарова, Е. А. Никифорова*

## МЕТОДИКА «ПОНЯТИЙНЫЙ СИНТЕЗ» КАК ИНСТРУМЕНТ ИНТЕЛЛЕКТУАЛЬНОЙ ДИАГНОСТИКИ: КАЧЕСТВЕННЫЙ АНАЛИЗ И СОДЕРЖАТЕЛЬНЫЕ КОММЕНТАРИИ

Методика «Понятийный синтез» была предложена М. А. Холодной в рамках онтологического подхода к исследованию интеллекта как альтернатива существующим инструментам измерения интеллектуальной успешности. Статья посвящена обсуждению и обобщению опыта работы с методикой, качественному анализу заданий и пересмотру оригинальных критериев оценки ответов. На основании материалов, собранных в трех различных исследованиях (с 2013 по 2015 г.; всего 104 испытуемых) была составлена база данных, включающая в себя 924 ответа на задания методики «Понятийный синтез», оцененных согласно оригинальным критериям. Все ответы были повторно и независимо оценены тремя экспертами-психологами. Результатом такой работы стали модифицированные критерии оценки ответов. Также в статье представлен качественный анализ более 40 различных ответов. Обсуждаются типичные методические затруднения, возникающие при работе с оценкой заданий, и предлагаются возможные варианты их разрешения. Библиогр. 19 назв. Табл. 1.

*Ключевые слова:* интеллект, диагностика интеллекта, концептуальные способности, методика «Понятийный синтез».

*Щербакова Ольга Владимировна* — кандидат психологических наук, Санкт-Петербургский государственный университет, Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9; o.shcherbakova@spbu.ru

*Макарова Дарья Николаевна* — аспирант, Санкт-Петербургский государственный университет, Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9; d.makarova23@gmail.com

*Никифорова Елизавета Андреевна* — студент, Санкт-Петербургский государственный университет, Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9; elise-nikiforova@yandex.ru

*Shcherbakova Olga V.* — PhD, Saint Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation; o.shcherbakova@spbu.ru

*Makarova Daria N.* — Postgraduate, Saint Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation; d.makarova23@gmail.com

*Nikiforova Elizaveta A.* — Undergraduate, Saint Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation; elise-nikiforova@yandex.ru

# Introduction

Intelligence research has always been a controversial area of knowledge since it first appeared in the late XIX century [1; 2]. Due to the major disagreements on theoretical approaches to intelligence studies and even on the very concept of intelligence [3; 4], it is not surprising that up to the present day there are no unquestionable psychometric tools for intelligence measurement. Nevertheless, researchers and practicing psychologists are desperate to use some of them.

Traditional IQ tests are widely used and still remain the main diagnostic tool when it comes to intelligence measurement. But they are also well known for various flaws coming from both particular tasks' contents and general theoretical background [5; 6; 7]. J. Raven's "Standard Progressive Matrices" test (SPMT) is considered to be one of the most accurate and reliable tools for measuring intelligence (more specifically, fluid intelligence as a component of "general cognitive ability") [8; 9]. However, J. Raven's test is also often criticized for the lack of correspondence between SPMT scores and real-life intellectual achievements [10–13].

Alternative tools for measuring intellectual abilities were suggested by M. A. Kholodnaya as a development of an onthological approach to intelligence studies [14; 15]. These tests aim to measure categorical and conceptual abilities that underlie intellectual productivity. Although scores obtained in these tests do not seem to correlate highly with SPMT scores, there is an evidence for their correlation with successful solving of more "ecological" cognitive tasks [12; 13]. "Concept Synthesis" test is one of these alternative tools and the most powerful, too. It implies the necessity to derive a new conceptual knowledge by combining three words from completely different semantic contexts into one meaningful sentence. Such kind of a task might bear a faint resemblance to S. Mednick's "Remote Associates Test" [16]; however, "Concept Synthesis" tasks do not explicitly encourage participants to produce unusual or creative answers. Instead of that, its goal is to enhance a cognitive effort that will lead to emerging of new conceptual context that will embrace different pieces of information and provide them with objective semantic interconnections.

"Concept Synthesis" test is becoming more and more popular [10–15; 17–19]. In the meantime, our experience of using this research tool revealed some technical issues that needed to be discussed. In particular, we found out that the diversity of answers produced by the participants was so large, that in many cases original criteria suggested by M. A. Kholodnaya were hardly applicable. Also, we faced the fact that some answers, that, according to the existing criteria, were supposed to be rated the same way, were not equal in terms of their real cognitive "value": they looked the same on the surface, but were actually based on distinctly different cognitive operations requiring different mental efforts. These observations led us to an idea that original criteria might need to be updated. Thus, the aim of this study was to put together the results obtained in numerous studies that used "Concept Synthesis" test, to reflect on our experience of using this research tool and provide some conceptual commentaries on the original criteria.

# Materials and methods

We used the original version of **M. A. Kholodnaya's test "Concept Synthesis"** [14]. It was designed to measure the ability to devise a set of semantic contexts based on three

words from remote semantic fields. The participant is presented with 3 concrete words belonging to 3 different semantic fields. The task is to find out as much sense bearing semantic links between these words as possible and to write them down in 1 or 2 sentences using all three words. There are 4 word triads in total. The time limit for each three words is 3 minutes. The word triads are as following:

a) cockle-shell — paper clip — thermometer;
b) computer — tornado — pin;
c) planet — electrical socket — sandglass;
d) wheel — lightning — ruler.

For each triad, the answers are rated 0, 1, 2 and 3 scores depending on how concrete/generalized the semantic link between all three words is. M. A. Kholodnaya [14] suggests the following criteria for differentiating and scoring the answers:

0 — there are only two out of three target words in the answer:

**e.g. "A pin and a tornado could hurt a human";**

1 — all three target words are included in the answer, but either two of them are used in the same functional role being opposed to the third one or all three words are used in the same functional role:

**e.g. "A tornado lifted pins and a computer into the air";**

2 — all three target words are included in a description of a concrete situation:

**e.g. "Children were watching a movie about a tornado on a computer and playing with pins";**

3 — all three target words are linked together either on the basis of a generalized category or analogy, or by causal relation, or by common complex context:

**e.g. "A pin and a computer are artificial objects made by human, and a tornado is a natural phenomenon which is beyond human control" or "A pin got inside a computer, caused a short circuit, which made a man switch his attention from the computer and he heard the news about oncoming tornado".**

The sum of the scores for all 4 triads represents one's ability to create new semantic contexts basing on concrete words from remote semantic fields.

To figure out what answers were the hardest to assess and caused most doubts in scorers, we compiled a database of 924 answers for "Concept Synthesis" test, collected from 2013 to 2015 in three different studies (all together, 104 participants), and rated them in accordance with the original criteria. The distribution of the number of answers between triads and years is presented in the Table 1.

*Table 1.* **The distribution of the number of answers between triads and years**

| Year<br>**Triad** | 2013 | 2014 | 2015 | Overall |
|---|---|---|---|---|
| Cockle-shell — paper clip — thermometer | 32 | 123 | 79 | 234 |
| Computer — tornado — pin | 34 | 122 | 88 | 244 |
| Planet — electrical socket — sandglass | 27 | 116 | 70 | 213 |
| Wheel — lightning — ruler | 28 | 126 | 79 | 233 |
| **Overall** | 121 | 487 | 316 | 924 |

We implemented a non-standard scoring procedure while working with this database. In our previous studies [10; 13; 17], there usually were two or three experts, who rated the answers in cooperation with one another. As a result, there was one score for each answer that reflected experts' group judgment. This time we decided to ask three experts to rate answers independently. All three experts were females (mean age — 26 y.o.) and had an extensive prior experience of working with the "Concept Synthesis" test of 2 to 4 years. When the rating itself was completed, the experts had to compare their scores and discuss those that were different. After doing that, they had to reach a consensus by changing the criteria if necessary in the process of this discussion. These alternative and updated criteria and some of their conceptual commentaries on them became the main result of this work.

## Results and discussion

We suggest the following updates for original criteria, resulting from a detailed qualitative analysis of each answer, that was not scored in the same way by all the three experts. To explain how and why the criteria were changed, some of the typical answers, which were the most puzzling and, thus, led to criteria changing are presented below.

One of the first problems the experts encountered was the following: though some answers were formally correct (they were full, grammatically coherent sentences and had all three target words used) and should be scored "1" in according with the original criteria, they were, in actual fact, meaningless and sometimes even absurd. For example:

1. *Take a rope out of a thermometer, attach a paper clip and catch a cockle-shell.*
2. *A planet was getting energy from an electrical socket: its time was running out as sand in a sandglass.*
3. *Sandglasses are vials; you'll look through a glass jar at the planets in the light from a lamp plugged in an electrical socket.*
4. *Time in a sandglass is independent of both planet and electricity in an electrical socket.*
5. *I was riding a wheel and was measuring a snail, but a lightning struck me and vaporized the ruler.*

In all these answers, three target words are put together with no regards to their conceptual or categorical characteristics, almost randomly. Although answers of this kind require some mental fluency (to associate target words with one another, like in (**1**) or (**3**)) and formal language skills (to build up a linguistically correct sentence), they do not involve any real thinking processes, as they do not reveal any hidden relations between three target words. These answers might look as if there is some implied broader context (like in (**2**) or in (**5**)) that could explain and justify their "oddity"; in such case they could be even considered creative. But, since this context is not explicated in the answer itself, it should be viewed as pseudocreativity or, in other words, the lack of productive thinking masked with an odd imagery derived from the wording chosen.

Another group of answers included sentences, which were not absurd, but were meaningless in terms of the given task: none of the three target words meanings were considered while devising these sentences, so the target words could also be replaced by almost any other words.

For instance:

6. *A cockle-shell, a paper clip and a thermometer are useless things for fixing a toilet.*
7. *The task was to associate three completely different things: a cockle-shell, a paper clip and a thermometer.*
8. *A computer, a tornado and a pin are three completely different things.*
9. *When a tornado is raging around, neither a pin nor a computer are of help.*
10. *An electrical socket as well as sandglass and everything else on the planet belongs to our planet.*
11. *An absolutely ridiculous combination of a wheel, a ruler and a lightning.*

Answers of this group look more "normal": they are not as odd as the answers from the previous group and many of them are formally meaningful statements as they refer to an objective reality. Their weak point is that they are predominantly superficial, self-evident and even banal (**8**). They either involve some random context (**6**) or refer to the very necessity of the task solving (**7**) that leads to a superficial and formal answers, masking the absence of productive thinking. In both cases, this might be viewed as a lack of cognitive contact with the task or low motivation. To compose such answers, one does not need to make any cognitive effort that would highlight the essential characteristics of the target words and link them together on the grounds of their conceptual features. Thus, their cognitive value is extremely minor.

Another subgroup of answers of this kind consisted of sentences, which also involved all three target words, and one or two of them might even have been used with some consideration for their meanings. But the sentences as a whole were meaningless since they were nothing but artificial combinations of words which did not bear any real sense and did not really provide the reader with any new information. For example:

12. *Her wonderful ribbon attached with a pin was torn off during a tornado — what would she do without an advice from a computer.*
13. *Make a model of a planet with lights, plug a battery in an electrical socket, charge the battery and connect it to the model — and see how long it will last.*
14. *You can draw both lightning and wheel using a ruler.*

According to the existing criteria, all of the above answers were supposed to be rated not lower than "1", because only answers, which included two out of three target words, were supposed to be rated "0". But the experts decided that all of these answers (and all answers of the same type) should also be rated "0" since they are not based on any conceptual or categorical relations between the target words and do not require any real mental effort. Thus, the criterion for the answers, that should be scored "0", was updated.

The second issue the experts encountered was the following: though many answers were meaningful sentences involving all three target words and describing a realistic situation — so all of them were supposed to be scored "2" — devising of some of those answers required simpler mental operations, than devising of others. As a result, there was no consistency within the answers that were scored "2" as some of them were of a higher quality than the others. The main characteristic of one group of the "lower quality" answers was that one of the target words was used with quotation marks (as a name for something). For instance:

15. *I bound the lists of my story called "Cockle-shell" with a paper clip but when I started to get ready, I hit a thermometer.*

16. *A revision in a club called "Planet" showed that a sandglass was broken and an electrical socket wasn't working.*

Practically, the use of this cognitive strategy simplifies the task by excluding one of the target words from real conceptual processing. Being enclosed with quotation marks and, thus, put aside the main narration line, a word loses its initial categorical and conceptual features and can be used as a label for almost anything. We tend to view this strategy as a way to make the task less difficult (either purposefully or not) and save mental resources. Although it might look to some extent similar to other strategies (e.g. using only two target words out of three) described above, we consider this one to be gentler and more elaborated.

The main characteristic of the second group of the "lower quality" answers was that the context devised in these sentences was quite plain, with not too many details. For example:

17. *A cockle-shell was attached to a thermometer with a paper clip.*
18. *I printed an image of a tornado using a computer and attached it to wallpaper with a pin.*
19. *A pin brought in by a tornado can damage a computer.*
20. *A sandglass with images of our planet on both sides stood not far from an electrical socket.*
21. *A lightning struck a wheel — the traces of a lightning could be measured with a ruler.*
22. *I came up to a car and kicked a wheel holding a ruler in my hand, suddenly a lightning flashed.*

These answers seem to combine three target words either by including them in a very simple visual scene based on the elements of common everyday experience (**17**, **18**) or by lining up such scenes into a sequence (**21**, **22**).

According to the existing criteria, all of the above answers were supposed to be rated "2" as they described a concrete situation and none of the three target words were used in the same functional role. But the experts decided that all of these answers (and all answers of the same type) should be rated "1" since there was a lack of conceptual relations between all the three target words. Thus, the criterion for the answers that should be scored "1" was also updated.

Other answers that also described a concrete situation were still scored "2" because the context devised in them was much more complex than in previous examples (which were scored "1"). Some of those answers are presented below:

23. *Picking up a paper clip that fell from the table, the doctor saw a cockle-shell lying on the bookshelf, which reminded him of his last holiday at the seaside; after that, getting back to work, the doctor asked a patient to give a thermometer back so he could check patient's temperature.*
24. *I was collecting cockle-shells in the cold sea, so now I'm lying in my bed with a thermometer, unbending a paper clip.*
25. *The cat raced through the apartment like a tornado after he had accidentally got pricked with a pin, and pushed a computer off the table.*
26. *I watched the news on my computer and found out that a tornado occurred in America; I marked a place where it was last seen on the map with a pin.*

27. *In the laboratory there was even a model of a planet, which started to spin when plugged in an electrical socket; also, there was an enormous sandglass dated late XVIII century.*
28. *Apocalypses is not a very likely event to happen, I don't think that it's possible that all electrical sockets will stop working, so please take this sandglass to an antique shop.*
29. *Working on the blueprints with a ruler, the architect saw a lightning through the window, which was coupled with a sound of thunder; because of that, he didn't hear the sound of wheels of a car moving on the gravel near the house.*
30. *While they were measuring the height of an old garage with a ruler, Peter had to stand on an old wheel and it was then when he saw a beautiful lightning.*

According to the existing criteria, all of the above answers were supposed to be scored "2", because they described a concrete situation — and so they were. Therefore, the experts had to update the criterion for answers, which should be scored "2", in order to make it possible to tell the difference between the 2 groups of answers describing a concrete situation (sentences with a plain context and sentences with a more complex, enriched context).

Most of the answers that experts scored "3" were highly creative. They were based on rich and complicated multi-component context that was elaborated by participants on their own initiative, without any instructions to do so, and involved many details that were not derived directly from the original task (**31, 33, 36, 38, 39**); they also involved remote analogies (**35, 42**), humor (**37, 40, 42**), metaphors (**32, 38**) and eloquent expressions (**34, 41, 42**), since target words were sometimes used in figurative meaning. Importantly, these answers were composed by playing upon unobvious (although still essential) characteristics of the target words that led to the evolving of new smart and meaningful combinations (**37, 40**). Finally, these answers were efficient as they were highly coherent and accomplished.

31. *A thermometer, which hung on the wall, had cockle-shells forming small boats on it, and a paper clip was a fishing rod of a sailor who was sitting on the bank.*
32. *It was 20 degrees on a thermometer, so I finally got out of my cockle-shell and went to the store for paper clips and pencils.*
33. *I put a cockle-shell to my ear and started to imagine that I was on vacation at the seaside, walking on the sand, stumbling on garbage buried by other tourists: cigarette buds, candy wrappers, paper clips; a doctor took a thermometer out. I'm not going to the sea this year.*
34. *A tornado raged through Silicon Valley, and it's easier to find a pin in a haystack than a working computer.*
35. *These days information can be transmitted from one computer to another with the speed close to the speed of a tornado, one computer is connected to other computers with wires as pieces of cloth are attached together with a pin.*
36. *When a tornado is simulated on a computer with the use of sensors, special antennas, which look like pins, are used: they are thrown into a tornado and transmit radio signal.*
37. *She had an amazing sandglass figure, lush hair (as if she had put her fingers into an electrical socket), and it seemed that she herself came from some unknown planet.*

38. *Time was running out as sand in a sandglass, it seemed that the whole planet was laughing at me, so the only thing left was to put my fingers into an electrical socket and get it over with.*

39. *The sandglass showed that there were only few seconds before the planet destruction, and he forgot that the moment he unplugged the computer from an electrical socket the game would be over.*

40. *After a lightning struck my Bentley, even the wheels burned down, and I think that a one meter long ruler won't be long enough to measure the pile of money I lost.*

41. *A squirrel was running inside a wheel with a speed of lightning, but it suddenly tripped over a ruler that was shoved in between the bars of a cage.*

42. *The wheel of fortune had moved: a lightning struck a temple spire the moment the teacher lifted a ruler above my fingers.*

Thus, based on the qualitative analysis and experts' focus group discussion, we found it possible to slightly update the criterion for the answers that should be scored "3".

An updated version of the criteria is presented below.

0 — an answer involves only two out of three target words or the answer is meaningless (or absurd); the answer is a phrase, not a sentence; the answer is formally correct, but designed in such a way that the three target words could be replaced with almost any other words;

1 — all the three target words are included in a description of a concrete situation; either two of them are used in the same functional role being opposed to the third one or all the three words are used in the same functional role (regardless of their specific features);

2 — all the three target words are included in an enriched context regardless of their specific features, or at least two of them are included in the context by the use of their specific features;

3 — a highly creative answer, in which all the three target words are linked together by either causal relation or by a common complex context.

### Limitations and conclusions

The goal of this study was to discuss the research tool that is growing in popularity within cognitive and intelligence research. "Concept Synthesis" test is based on the onthological approach to intelligence studies [14; 15] and is supposed to measure productive intellectual abilities that make production of the new knowledge possible. This is why it was important to reflect on 4 years of extensive research work with the use of this test and provide some conceptual commentaries on it. The updated criteria for assessing the answers for "Concept Synthesis" tasks were suggested as a result of qualitative analysis carried out by 3 experts. These criteria might be useful for differentiating between similarly looking, but conceptually contrasting answers, that, in its turn, will be helpful for future advances in psychological diagnostics. However, this tentative attempt to develop the existing tool should not be viewed as a substitution for proper standardization. One of the next steps in advancing the research methodology for intelligence studies could be actual standardization based on an extended pool of answers. The qualitative commentaries that were presented in this paper can set the stage for it.

## Acknowledgements

## References

1. Spearman C. "General Intelligence", objectively determined and measured. *The American Journal of Psychology*, 1904, vol. 15, no. 2, pp. 201–292.
2. McNemar Q. Lost: Our Intelligence? Why? *American Psychologist*, 1964, vol. 19, no. 12, pp. 871–882.
3. Neisser U. The concept of intelligence. *Intelligence*, 1979, vol. 3, no. 3, pp. 217–227.
4. Legg S., Hutter M. A collection of definitions of intelligence. *Frontiers in Artificial Intelligence and Applications*, 2007, vol. 157, pp. 17–24.
5. Burks B. S., Jensen D. W., Terman L. M. The promise of youth. Ed. by L. M. Terman. *Genetic studies of genius* (Vol. 3). Stanford, CA, Stanford University Press, 1930.
6. Boring E. G. Intelligence as the Tests Test It. Eds. J. J. Jenkins, D. G. Paterson. *Studies in individual differences: The search for intelligence*. East Norwalk, CT, US, Appleton-Century-Crofts, 1961, pp. 210–214.
7. Kholodnaya M. A. *Psychology of Intelligence: Research Paradoxes*. 2$^{nd}$ ed., revised and enlarged. St. Petersburg, Piter, 2002. (In Russian)
8. Cattell R. B. *Abilities: Their growth, structure, and action*. Boston, Houghton Mifflin, 1971.
9. Carpenter P. A., Just M. A., Shell P. What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological review*, 1990, vol. 97, no. 3, pp. 404–431.
10. Makarova D. N., Shcherbakova O. V. Conceptual Abilities and Their Impact to Psychometric IQ and Expertise. *Abstracts of 16$^{th}$ Annual Conference of International Society for Intelligence Research (ISIR)*. Albuquerque, USA, September 18–20, 2015, p. 39.
11. Makarova D. N., Nikiforova E. A, Shcherbakova O. V. Is general intelligence not that general? Evidence from SPMT and conceptual abilities tests. *Abstracts of the 17$^{th}$ Annual Conference of International Society for Intelligence Research (ISIR)*. Saint Petersburg, Russia, July 15–17, 2016, p. 84.
12. Nikiforova E. A., Shcherbakova, O. V. Conceptual abilities as a predictor of apologues comprehension efficiency. *International Journal of Psychology. Special Issue: 31st International Congress of Psychology, 24–29 July 2016, Yokohama, Japan*. July 2016, vol. 51, issue Supplement S1, pp. 176–177.
13. Shcherbakova O. V., Makarova D. N. Are they different or the same? Experts' psychometric IQ does not correlate with conceptual abilities. *Vestnik SPbSU. Series 16. Psychology. Education*, 2016, issue 4, pp. 88–96.
14. Kholodnaya M. A. *Psychology of Conceptual Thinking: From Conceptual Structures to Conceptual Abilities*. Moscow, Institute of Psychology RAS Print, 2012. (In Russian)
15. Kholodnaya, M. A., Volkova, E. V. Conceptual Structures, Conceptual Abilities and Productivity of Cognitive Functioning: The Ontological Approach. *Procedia-Social and Behavioral Sciences*, 2016, vol. 217, pp. 914–922.
16. Mednick S. A., Mednick M. *Remote associates test: Examiner's manual*. Houghton Mifflin, 1971.
17. Shcherbakova O. V., Makarova D. N. Conceptual abilities as a basis for expertise. *15$^{th}$ Annual Conference of the International Society for Intelligence Research (ISIR)*. Graz, Austria, December 12–14, 2014, p. 62.
18. Sipovskaya Ya. A. Conceptual, metacognitive and intentional descriptors of intellectual competence in late adolescence. *Vestnik SPbSU. Series 12. Psychology. Sociology. Education*, 2015, issue 4, pp. 22–31. (In Russian)
19. Trifonova A. V. The development of conceptual abilities in late adolescence and early adulthood. *Akmeologiya*, 2015, no. 2 (54), pp. 144–148. (In Russian)