



Published in final edited form as:

J Creat Behav. 2020 March ; 54(1): 62–74. doi:10.1002/jocb.345.

Validation of the Italian Remote Associate Test

Carola Salvi^{1,2}, Giulio Costantini³, Adriana Pace⁴, Massimiliano Palmiero^{4,5}

¹Department of Psychology and Cognitive Brain Mapping Group, Northwestern University, Evanston, IL

²Rehabilitation Institute of Chicago, Illinois, USA

³Department of Psychology, University of Milano-Bicocca, Milan, Italy

⁴Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, Italy

⁵Neuropsychology Unit, I.R.C.C.S.Fondazione Santa Lucia, Rome, Italy

Abstract

The scientific approach to the study of creative problem solving has shifted from using classic insight problems (e.g., the *Nine-dots* problem), towards sets of problems that have more robust psychometric properties, such as the Remote Associate Test (RAT). Because it is homogeneous, compact, quickly solvable, and easy to score, the RAT has been used more frequently in recent creativity studies. We applied the Item Response Theory (IRT) to develop an Italian version of this task. The final 51-item test was reliable ($\alpha = .89$) and provided information over a wide range of ability levels, as revealed by the IRT analysis. The RAT correlated with five measures of creative performance: The Raven's Standard Progressive Matrices (SPM), three classic insight problems, a set of anagrams purposefully developed, the fluency and flexibility scores of the Alternative Uses Task (AUT), and the Creative Achievements Questionnaire (CAQ). The new measure provided is meant to encourage the study of creativity and problem solving in the Italian language.

Keywords

Problem Solving; Remotes Associates Task; Creativity; Convergent Thinking; Anagrams; Divergent Thinking

Based upon Guilford's approach (Guilford, 1967), creativity requires a combination of both divergent and convergent thinking. Divergent thinking aims to generate multiple answers for an open-ended problem, whereas convergent thinking targets a single correct solution for a closed-ended problem. Influenced by Guilford's approach, Mednick (1962) developed the associative theory of creativity. According to this theory, creative thinking consists in "the forming of associative elements into new combinations which either meet specified requirements or are in some way useful" (Mednick, 1962, p. 221). Therefore, according to Mednick, creativity involves accessing remote concepts that were previously unconnected, or dissimilar thought elements. This theory was operationalized using the Remote

Associates Test (RAT; Mednick & Mednick, 1967) as a measure of convergent thinking. In particular, the RAT consists of two sets of 30 triads of words (Mednick, 1968; Mednick & Mednick, 1967), where each word can be associated with a fourth word by creating a compound word, via semantic association, or synonymy. For instance, *same*, *tennis*, *head* can be associated with the word *match* by creating the compound word *match-head*, by semantic association (i.e., *tennis match*) and because of synonymy (i.e., *same = match*). Afterwards, the CRA, a measure similar to the RAT, have been developed, (Bowden and Jung-Beeman, 2003b; see Salvi, Costantini, Bricolo, Perugini, & Beeman, 2015 for the Italian version), where the solution word always forms a compound word or a common two-word phrase with each problem word (e.g., in the problem *crab*, *pine*, *sauce*, the solution *apple* forms the compounds *crabapple*, *pineapple*, and *apple sauce*). Thus, whereas the CRA is based upon a more consistent rule, the RAT is based upon a more flexible rule leading to a broader semantic search that involves association and synonymy in addition to word matching. Researchers used one set or the other according to experimental designs needs and procedures.

Before the introduction of the RAT and similar tests, creative problem solving was typically assessed using classic insight problems (e.g., Bowden, Jung-Beeman, Fleck, & Kounios, 2005; Dreistadt, 1969; Ohlsson, 1984). However, these problems were too few in number, non-homogenous or compact, and took too long to solve, to constitute a useful set of stimuli for many modern research paradigms; For example, those involving electroencephalography and neuroimaging (e.g., Bowden, Jung-Beeman, Fleck, & Kounios, 2005). Most importantly, recent research demonstrated that performance in classic insight problems does not predict real-world creativity (Beaty, Nusbaum, & Silvia, 2014). On the contrary, the RAT, and similar versions of it, have been successfully used in the study of problem solving, cognitive flexibility, and creative thinking (e.g., Ansburg, 2000; Bowden & Beeman, 1998; Bowden & Jung-Beeman, 2003a; Dominowski & Dallob, 1995; Jung-Beeman & Bowden, 2000; Salvi et al., 2015; Salvi, Bricolo, Franconeri, Kounios, & Beeman, 2016; Salvi, Cristofori, Beeman & Grafman 2016).

The RAT presents several benefits (Bowden et al., 2005):

- a. It is compact and easy to present item by item on a computer screen.
- b. It is easier to solve than classic insight problems.
- c. It includes items formed by single or few-words solutions and therefore can be used in experimental paradigms such as neuroimaging and solution priming.

Nevertheless, the RAT items are contextualized by the language. For instance, several compound words (e.g., *pine-apple*) and idioms (e.g., *lip service*) do not make sense in languages other than English. Therefore, the development of creative tests in different languages cannot be performed solely through literal translation. It is necessary to develop an entirely new set of problems, which needs to be grounded in the specific characteristics of the new language, in its vocabulary, and in its colloquial conventions. To our knowledge, English, Hebrew (Nevo & Levin, 1978), Jamaican (Hamilton, 1982), Japanese (Baba, 1982) and Dutch (Akbari Chermahini, Hickendorff, & Hommel, 2012) versions of the RAT have been developed, whereas no Italian version is available.

Therefore, the main aim of this study is to develop an Italian version of the RAT and provide validity evidence for it. To this end, we first created a preliminary pool of RAT problems using a pretest (for details see the Supplementary Material S1). Second, we selected a subset of problems according to their psychometric properties using the Item Response Theory (IRT). Third, we assessed the construct validity of the newly developed test (Campbell & Fiske, 1959; Hair, Black, Babin, Anderson, & Tatham, 2010) with respect to measures of convergent and divergent thinking as well as creative achievements.

The RAT is a measure of convergent thinking. Therefore, to evaluate its convergent validity, in line with Akbari Chermahini et al. (2012), we also administered insight problems that target convergent thinking (e.g., Ansburg, 2000), and share with the RAT initial difficulty in conceptualizing the solution and restructuring (Cunningham, McGregor, Gibb & Haar, 2009). We also administered Raven's matrices and purposely-developed anagrams, which also represent aspects of convergent-thinking performance. Significant correlations were expected between the RAT and the just mentioned measures of convergent thinking.

We also administered the Alternative Uses Task (AUT - Guilford, 1967), an established measure of divergent thinking. Solving the RAT might require both generating novel candidate responses (e.g., a number of words that are associated to each of the words of a RAT item) and quickly exploring each alternative to accept it as the correct response or to reject it. This means that besides convergent thinking the RAT also involves the ability to find relationships between elements and fluency of thought (e.g., Ansburg, 2000). In this vein, although not systematically (e.g., see Colzato, de Haan, & Hommel, 2015), different studies found correlations between the RAT and measures of divergent thinking (e.g., Jones & Estes, 2015; Taft & Rossiter, 1966). Therefore, at least few correlations between the RAT and components of the AUT, such as fluency (generation of novel alternatives) were expected.

Yet, in order to further elucidate the validity of the RAT, we administered the Creative Achievements Questionnaire (CAQ - Carson, Peterson, & Higgins, 2005), a self-reported measure of the effects of creative thinking in everyday life. Since creative achievement is based on different cognitive processes including divergent and convergent thinking, positive correlations between the RAT and CAQ were also expected (e.g., Piffer, 2014 found that the RAT was only related to the scientific achievement score); however, literature in this regard is scarce.

Method

Participants

One-hundred and fifty-three Italian students (M Age = 24.8; SD = 4.9; 95 females) from the University of L'Aquila participated in the study for partial course credit. All participants were right-handed Italian native speakers. By the anamnesis questionnaire, they were healthy and without neurological and/or psychiatric disorders; they did not report alcohol or drug addiction. The study was designed in accordance with the ethical principles of human experimentation stated in the Declaration of Helsinki, and was approved by the Institutional

Review Board of the Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila.

Materials

Italian RAT

An Initial set of seventy-eight RAT problems was developed by means of a pretest (Pretest 1; see Supplementary Material for details). RAT problems were randomly presented on a computer screen and preceded by four practice trials¹. Participants were instructed to press the space bar as soon as they found a solution and to enter the solution word in the subsequent screen. No feedback was given regarding the accuracy of the solution. Ability estimates were computed using the expected a posteriori method (Kolen & Tong, 2010; see section Analyses for details).

Italian Anagrams

Participants completed sixty-five anagrams presented on a computer screen in randomized order, following the same procedure used for the RAT. Anagrams were preceded by three practice trials. Since the Italian Anagrams were also a newly developed test, we performed an IRT analysis on Anagrams as well, which resulted in a final 45-item test (details of this analysis are reported in the Supplementary Material). Ability estimates were computed using the expected a posteriori method (Kolen & Tong, 2010).

Classic insight problems

Three classic insight problems covering three different domains were used. The original English version of the verbal insight problem was translated into Italian and then an English-speaking translator checked it for correctness. Instructions were also adapted to Italian context. Insight in the mathematical domain was assessed using the *Parallelogram problem* (e.g., Ohlsson, 1984; Segal, 2004). Participants were asked to compute separately the area of the square and that of the parallelogram (See Figure 1).

Insight in the linguistic domain was assessed by the *Bald man problem* (Bowden, 1985; Schwert, 2007): “A man was caught in the rain with no hat or umbrella. There was nothing over his head and his clothes got soaked, but not a hair on his head got wet. How is this possible?” *Solution*: the man was bald. Insight in the visuo-spatial domain was assessed by the *Star problem* (Dreistadt, 1969), which consisted in figuring out how to plant ten trees into five rows. *Solution*: rows must be arranged into a five-pointed star (See Figure 2). An overall score, ranging from 1 to 3, was computed as the number of problems solved.

¹The specific instructions for the RAT problems were: “For each problem you are presented with 3 words. Your task is to find a fourth word that can be associated with the three words displayed, in order to build three different associations: by synonymous (find which one of the three words can be combined with your solution word in order to get a synonymous); by semantic association (find which one of the three words can be combined with your solution word in order to get a semantic association); by forming a compound word (find which one of the three words can be combined with your solution word in order to get a compound word). For each problem you have up to 30 seconds to find the solution. As soon as the solution prompts in your mind press the space bar; please do not press the space bar if you do not have the solution, or if you are not sure about your solution. Immediately after having pressed the space bar, you must declare your solution). If you do not find the solution within 30 seconds, another problem is presented. The practice trials were: Mento-Scossa-Elettrica (Scarica); Mento-Accesso-Vetri (Porta); Otto-Pattuglia-Disco (Volante); Moto-Linea-Via (Retta).

Raven's Standard Progressive Matrices (SPM; Raven, 1965; Italian validation by Spinnler & Tognoni, 1987)

Raven's (1965) version of 5 series of 12 items each was administered. Participants were instructed to identify from a set of alternatives the missing piece of a visual pattern. Items became progressively more difficult. An overall score was computed as the number of problems solved.

Alternative Uses Task (AUT - Guilford, 1967; see also Palmiero, 2015; Palmiero et al., 2017)

Participants were asked to list as many possible uses as they could for the common object *brick* within 10 min. Scoring included three components: *Fluency*, the total number of relevant responses; *flexibility*, the number of different categories that were encompassed in the responses; and *originality*, that was computed following Runco, Okuda, and Thurstone (1987)² and which has been successfully used in several studies (e.g., Palmiero, 2015, Palmiero et al., 2010, 2017).

Creative Achievement Questionnaire (CAQ – Carson et al., 2005)

The CAQ investigates creative achievements in 10 domains (visual arts, music, dance, individual sports, team sports, architectural design, entrepreneurial ventures, creative writing, humor, inventions, scientific inquiry, theater and film, and culinary arts). For each domain, participants were asked to provide an estimate of their talent, ability, or training by check-marking items that applied to them. The scores assigned to items were different (e.g., "I have no training or recognized talent in this area" received 0 points, "My architectural design has been recognized in a local publication" received 6 points). Some items required participants to indicate the number of times a sentence applied to them (e.g., "I have received a patent for one of my inventions") and received a score that depended both on item and on frequency it applied to the participant. A total creative achievement score was computed by summing the number of points collected within each area (for full details on the scoring procedure, see Carson, Peterson, & Higgins, 2005).

Procedure

Participants were tested individually and were given a time limit of 30s for each RAT and Anagram trial, 5 minutes to solve each classic insight problem, 1 minute for each Raven Progressive Matrices and 10 minutes to solve the AUT, For the CAQ no time limit was used. The order of presentation of tests was randomized.

Analyses

We evaluated the psychometric properties of the RAT test using IRT (for a comprehensive introduction to IRT, see De Ayala, 2009). IRT models the probability of a correct response to each item as a function of a common latent variable (θ), representing the ability underlying test performance, which is assumed to follow a standard normal distribution. This ability has

²Each response was evaluated according to the relative frequency in the sample composed of N participants; responses were scored between $1/N$ (only when one participant proposed the idea), and $N/N=1$ (when all participants came up with the same idea). Then the score was subtracted from 1. The sum of these scores across responses was used as the individual originality score.

been described as creative convergent thinking in the case of the RAT (Akbari Chermahini et al., 2012). According to the IRT model, the probability of a correct response to each item increases with θ following a sigmoid curve, bounded between 0 and 1. The sigmoid curve is called the *Item Characteristic Curve* and can be parametrized with up to three parameters, which describe the properties of each item: *difficulty*, *discrimination*, and *guessing*. Item difficulty is the level of θ corresponding to a 50% probability of a correct response: Everything else being equal, more ability θ will be required to solve a difficult item. Item *discrimination* regulates the slope of the curve: An item with a high discrimination is generally better at discriminating among individuals that have a different stance on θ than an item with lower discrimination. The *guessing* parameter is the lower asymptote of the curve and represents the probability of selecting the correct answer simply by guessing, independent of θ . The IRT model that includes all these three parameters is the three-parameter logistic (3PL) model. This model is useful for modeling multiple-choice items, which allow guessing the correct response. In the case of closed-ended questions (like the RAT) one can assume no chance of responding correctly by guessing: the model in which the guessing parameter is constrained to be zero for all items is the two-parameter logistic (2PL) model. The one-parameter logistic (1PL) model assumes that items can vary in difficulty, but have the same discrimination. IRT allows computing the *Item Information Curve*, the amount of information conveyed by each item at each level of θ . For instance, everything else being equal, a very easy item will be more useful for assessing participants with low levels of θ , whereas a difficult item will be more useful for assessing participants with elevated levels of θ . The *Test Information Curve* is simply the sum of all item information curves and indicates the amount of information conveyed by the test at each level of θ . Participants' abilities and items' difficulties share the same metric, and therefore can be visualized together: The *Wright map* (Wilson, 2005) represents an histogram of a person's abilities in the left-hand side of the graph and the difficulty of each item in the right-hand side, on the same logit scale. A person whose ability estimate is above an item's difficulty is more than 50% likely to solve the item, whereas a person whose ability estimate is below an item's difficulty is less than 50% likely to solve that item.

We performed an IRT analysis in R (R Core Team, 2017) using packages *ltm* (Rizopoulos, 2006), *mirt* (Chalmers, 2012), and *Wright Map* (Irribarra & Freund, 2014). Given the open-ended response format of the RAT, we did not consider a 3PL model. Instead, we selected the model that provided the most parsimonious account of the data between the 1PL and the 2PL, according to the AIC (Akaike, 1974) and the BIC (Schwarz, 1978) criteria. IRT models assume unidimensionality, which means that all items assess the same underlying ability θ . We evaluated the unidimensionality assumption using the modified parallel analysis procedure proposed by Drasgow and Lissak (1983), which compares the second eigenvalue of the tetrachoric correlation matrix to the same eigenvalue in 100 datasets simulated under the same IRT model and returns a bootstrap-based p-value (a p value < .05 indicates a significant violation of the unidimensionality assumption).

We computed item-fit statistics, which indicate how well the observed pattern of responses to each item is predicted by the model (Ames & Penfield, 2015). We considered three item-fit statistics: outfit, infit (Wright & Panchapakesan, 1969), and the scaled version of the chi-square statistic (Stone, 2000). Outfit weighs equally deviances from predicted values,

whereas infit assigns a higher weights to individuals whose ability is close to the difficulty of the item, values of infit and outfit close to one indicate good fit, values between .5 and 1.5 are considered acceptable (Ames & Penfield, 2015). Stone's (2000) statistic employs a Monte Carlo method to account for the uncertainty around ability estimates. Infit and outfit are associated to z-values whereas Stone's statistic is associated to a p -value, z-values larger than 2 and p -value lower than .05 indicate a rejection of the null hypothesis that an item's data are consistent with the model. We iteratively removed items with the largest misfit, until there were no misfit items. We computed RAT ability estimates using the expected a posteriori method (Kolen & Tong, 2010) as implemented in the *mirt* package (Chalmers, 2012) and used these scores to evaluate the correlations between the RAT and the other measures administered.

Results

We examined how both 1PL and 2PL IRT models fit the initial 78-item RAT. Although the likelihood ratio test indicated that 2PL model fit the data better than the 1PL model, $LRT(77) = 179.753$, $p < .001$, the improvement in fit obtained by considering the 2PL model instead of the 1PL model (computed as suggested by De Ayala, 2009, p. 141) was only 1.69%. The AIC-values were in favor of the 2PL model, whereas the BIC pointed to the 1PL model as the most parsimonious. We computed each participant's ability estimates both considering the 1PL and the 2PL model: The Pearson's correlation between these estimates was nearly perfect ($r = .99$, $p < .001$), indicating that considering the 1PL or the 2PL model did not make a practical difference for estimating individual abilities. Therefore, we preferred the 1PL model, because it provided an account of the data as good as the 2PL model using around half parameters. The modified parallel analysis indicated that the unidimensionality assumption was not significantly violated ($p = .06$). The IRT parameters and fit statistics for the initial 78-item RAT are reported in the Supplementary Table S1, the corresponding item characteristic curves, test information function and Wright map are reported in Figure S1.

Some items showed significant misfit to the model $p < .05$. We iteratively excluded misfit items from the test and re-fitted the model, until only items that fit the model remained in the set. Twenty-seven RAT items were excluded in this process, resulting in the final 51-item test. The fit statistics of the 1PL and 2PL on the final set of items are reported in Table 1, the likelihood ratio test indicates that choosing the more complex 2PL models does not improve fit over the 1PL model, $LRT(50) = 64.44$, $p = .081$, and both AIC and BIC indicate the 1PL model as the most parsimonious. In the final item set, the modified parallel analysis indicated that the unidimensionality assumption was not violated ($p = .22$).

Psychometric properties of the final item sets

The final set of 51 RAT items is reported in Table 2, together with the IRT parameters and the fit statistics for each item. The item characteristic curves, the test information function and the Wright map are reported in Figure 3. The item difficulties took a wide range of values, ranging between -5.63 (extremely easy item) and 4.19 (very hard item; see Table 3 and Figure 3). This is reflected in the Wright map and in the Test Information Function, that

show that the RAT provided information for a broad range of ability levels, with a peak around zero, corresponding to typical ability levels.

On average, participants solved 27.46 RAT problems out of 51 ($min = 8$, $max = 46$, $Mdn = 28$, $SD = 8.35$) and 26.40 Anagrams out of 45 ($min = 4$, $max = 45$, $Mdn = 25$, $SD = 9.05$). Table 3 reports descriptive statistics and correlations among all measures administered. RAT and Anagrams were scored using the expected a posteriori ability estimates (Kolen & Tong, 2010). It is worth noticing that ability estimates showed sizable correlations with the simple count of correct responses for each participant ($r = .998$ for the RAT and $r = .989$ for the Anagrams, $ps < .001$). The classic insight problems and the CAQ showed unsatisfactory internal consistency. Some measures deviated from a normal distribution, as indicated by elevated kurtosis, therefore we focused primarily on the Spearman rank correlation coefficient (See Table 2), which does not assume a normal distribution and is more robust than Pearson's correlation to the presence of outliers (Wilcox, 2012). In detail, the RAT correlated significantly with the Anagrams, the Raven's SPM, the classic insight problems, and the general CAQ score. Given the low internal consistency of the classic insight problems, we further investigated the relation between the RAT and each of these problems separately. The RAT correlated significantly with the Parallelogram problem ($r_s = .31$, $p < .001$; $r = .31$, $p < .001$), the Bald-man problem ($r_s = .17$, $p = .036$; $r = .16$, $p = .048$), and the star problem ($r_s = .32$, $p < .001$; $r = .35$, $p < .001$). The RAT also correlated with the flexibility and fluency scores of the AUT, but not with the originality score. All AUT scales correlated significantly or marginally significantly also with the classic insight problems score.

Discussion

The aim of this study was to develop an Italian version of the RAT to expand the study of convergent thinking and insight problem solving to the Italian language, that is, to people that are proficient in Italian (63.4 Million L1 speakers; Lewis, Simons, & Fenning, 2016).

Using IRT analyses (see Table 2) we developed a 51-item RAT. Models based upon IRT are extensively studied and broadly adopted for investigating properties of tests and items. The use of IRT to develop the Italian RAT measure improved its psychometric properties and enhanced its validity. Indeed, as expected, the RAT conveyed information over a wide range of ability levels and it converged with classic insight problems, with the anagrams and with the SPM. The validity of the RAT was further supported by a significant correlation with the self-reported creative achievements, as measured by the CAQ, although the correlation between RAT and CAQ was lower than the correlations between RAT and other measures, indicating that RAT and CAQ share a relatively small portion of variance.

Interestingly, the RAT showed significant correlations with the AUT fluency and flexibility scores, but not with the originality score. This result confirms that the RAT, which is based on the ability to search for different words in the semantic network, is dependent upon some components of divergent thinking. In other words, "While the RAT clearly involves more top-down constraints on the cognitive search process than the AUT, it has a search component that requires flexibly moving from one memory trace to the next; and while the

AUT clearly involves more extensive and less constrained cognitive search than the RAT, it still involves some constraints” (Colzato et al., 2015, p. 710).

The correlations between the RAT and the AUT fluency and flexibility might be seen also as indicators of a lack of discriminant validity of the RAT as a pure measure of convergent thinking. However, one should consider that the AUT fluency and flexibility also correlated significantly with the anagrams and the classic insight problems, and marginally significantly with the SPM. This result gives further support to the idea that the creative process may require a combination of divergent and convergent thinking (Colzato et al., 2015; Cropley, 2006), and that it might be difficult to purely measure one of these processes and not the other one. In addition, the RAT did not correlate with the originality score of the AUT. The lack of correlation between these measures could indicate that the need for original responses is less accentuated in the RAT than in the AUT. Conversely, a significant correlation between originality and the classic insight problems suggests that the need for originality could be more relevant in solving this kind of problem.

In conclusion, in this work we provided a new measure of creative convergent thinking in the Italian language and culture. The overall pattern of results suggests that the Italian RAT is a good measure of creative convergent thinking. The difficulty estimates could be particularly useful for researchers that are interested in defining subsets of items to administer. A limit of this study was considering a single task, albeit an established one, for assessing divergent thinking, namely the AUT. An important task for future research is to deepen the connections between the RAT and divergent thinking using a wider array of divergent thinking measures. Creativity is associated with different domains of knowledge (e.g., Batt et al., 2010, Boccia, et al., 2015; Palmiero et al., 2015; Palmiero et al., 2016) and both the RAT and the anagrams tap particularly into the linguistic realm. An interesting task for future research is further examining the relationships between creative convergent thinking in the linguistic domain and creativity in other domains. Finally, another limitation regards the restriction to Italian speakers. However, one should consider that the most of measures are initially developed and validated in English, and can be used by about 372 Million L1 English speakers (Lewis, Simons, & Fenning, 2016). Therefore, increasing the number of potential test takers across cultures appears to be an important challenge for future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank xxx and xxx for the helpful edits and insightful comments on the draft of this manuscript. This work was supported by NIH under grant No xxxxxx to xxx, and by X.X.X.X.X.X. Fondazione xxx xxx, xx, Italy.

References

Akaike H. 1974; A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 19:716–723.

- Akbari Chermahini S, Hickendorff M, Hommel B. 2012; Development and validity of a Dutch version of the Remote Associates Task: an item-response theory approach. *Thinking Skills and Creativity*. 7:177–186.
- Ames AJ, Penfield RD. 2015; An NCME instructional module on item-fit statistics for Item Response Theory models. *Educational Measurement: Issues and Practice*. 34:39–48.
- Ansburg PI. 2000; Individual differences in problem solving via insight. *Current Psychology*. 19:143–146.
- Baba Y. 1982; An analysis of creativity by means of the remote associates test for adults revised in Japanese (Jarar Form-A). *Japanese Journal of Psychology*. 52:330–336.
- Batt R, Palmiero M, Nakatani C, van Leeuwen C. 2010; Style and spectral power: processing of abstract and representational art in artists and non-artists. *Perception*. 39:1659–1671. [PubMed: 21425703]
- Beatty RE, Nusbaum EC, Silvia PJ. 2014; Does insight problem solving predict real-world creativity? *Psychology of Aesthetics, Creativity, and the Arts*. 8:287–292.
- Beeman MJ, Bowden EM. 2000; The right hemisphere maintains solution-related activation for yet-to-be-solved problems. *Memory & Cognition*. 28:1231–41. [PubMed: 11126944]
- Boccia M, Piccardi L, Palermo L, Nori R, Palmiero M. 2015; Where do bright ideas occur in our brain? Meta-analytic evidence from neuroimaging studies of domain-specific creativity. *Frontiers in Psychology*. 6:1195. [PubMed: 26322002]
- Bowden E, Beeman MJ. 1998; Getting the right Idea: semantic activation in the right hemisphere may help solve insight problems. *Psychological Science*. 9:435–440.
- Bowden E, Jung-Beeman M, Fleck J, Kounios J. 2005; New approaches to demystifying insight. *Trends in Cognitive Sciences*. 9:322–8. [PubMed: 15953756]
- Bowden EM. 1985; Accessing relevant information during problem solving: time constraints on search in the problem space. *Memory & Cognition*. 13:280–6. [PubMed: 4046829]
- Bowden EM, Jung-Beeman M. 2003a; Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychonomic Bulletin & Review*. 10:730–7. [PubMed: 14620371]
- Bowden EM, Jung-Beeman M. 2003b; Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers*. 35:634–9.
- Campbell DT, Fiske DW. 1959; Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 56:81–105. [PubMed: 13634291]
- Carson SH, Peterson JB, Higgins DM. 2005; Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal*. 14(1):37–50.
- Chalmers RP. 2012; mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*. 48:1–29.
- Colzato LS, de Haan AM, Hommel B. 2015; Food for creativity: tyrosine promotes deep thinking. *Psychological Research*. 79:709–714. [PubMed: 25257259]
- Cropley A. 2006; In praise of convergent thinking. *Creativity Research Journal*. 18:391–404.
- Cunningham JB, MacGregor JN, Gibb J, Harr J. 2009; Categories of insight and their correlates: An exploration of relationships among classic-type insight problems, rebus puzzles, remote associates and esoteric analogies. *Journal of Creative Behavior*. 43:262–280.
- De Ayala, RJ. *The theory and practice of item response theory*. New York: Guilford Press; 2009.
- Dominowski, RL, Dallob, P. Insight and problem solving. In: Sternberg, RJ, Davidson, JE, editors. *The nature of insight*. Cambridge, MA: MIT Press; 1995. 273–278.
- Drasgow F, Lissak RI. 1983; Modified parallel analysis: a procedure for examining the latent dimensionality of dichotomously scored item responses. *Journal of Applied Psychology*. 68:363–373.
- Dreistadt R. 1969; The use of analogies and incubation in obtaining insights in creative problem solving. *The Journal of Psychology*. 71:159–175.
- Guilford, JP. *The nature of human intelligence*. Vol. 5. McGraw Hill; 1967.
- Hair, JF, Black, WC, Babin, BJ, Anderson, RE, Tatham, RL. *Multivariate data analysis*. 7th. Upper Saddle River, NJ: Prentice Hall; 2010.

- Hamilton MA. 1982; “Jamaicanizing” the Mednick Remote Associates Test of creativity. *Perceptual and Motor Skills*. 55:321–322.
- Jones LL, Estes Z. 2015; Convergent and divergent thinking in verbal analogy. *Thinking and Reasoning*. 21(4):473–500.
- Kolen MJ, Tong Y. 2010; Psychometric properties of IRT proficiency estimates. *Educational Measurement: Issues and Practice*. 29:8–14.
- Iribarra, DT; Freund, R. Wright Map: IRT item-person map with ConQuest integration. 2014. R package version 1.0. Retrieved from <http://github.com/david-ti/wrightmap>
- Lewis, MP, Simons, GF, Fennig, CD. *Ethnologue: languages of the world*. Vol. 16. Dallas, TX: SIL international; 2009.
- Mednick SA. 1962; The associative basis of the creative process. *Psychological Review*. 69:220–232. [PubMed: 14472013]
- Mednick SA. 1968; Remote Associates Test. *Journal of Creative Behavior*. 2:213–214.
- Mednick, SA, Mednick, MP. *Examiner’s manual: Remote Associates Test*. Boston: H. Mifflin, Ed; 1967.
- Nevo B, Levin I. 1978; Remote Associates Test: assessment of creativity in Hebrew. *Megamot*. 24:87–98.
- Ohlsson S. 1984; Restructuring revisited. An information processing theory of restructuring insight. *Scandinavian Journal of Psychology*. 25:117–129.
- Palmiero M. 2015; The effects of age on divergent thinking and creative objects production: a cross-sectional study. *High Ability Studies*. 26:93–104.
- Palmiero M, Nakatani C, Raver D, Olivetti Belardinelli M, van Leeuwen C. 2010; Abilities within and across visual and verbal domains: how specific is their influence on creativity? *Creativity Research Journal*. 22:369–377.
- Palmiero M, Nakatani C, van Leeuwen C. 2017; Visual creativity across cultures: a comparison between Italians and Japanese. *Creativity Research Journal*. 29:86–90.
- Palmiero M, Nori R, Aloisi V, Ferrara M, Piccardi L. 2015; Domain-specificity of creativity: a study on the relationship between visual creativity and visual mental imagery. *Frontiers in Psychology*. 6:1870. [PubMed: 26648904]
- Palmiero M, Nori R, Piccardi L. 2016; Visualizer cognitive style enhances visual creativity. *Neuroscience Letters*. 615:98–101. [PubMed: 26806864]
- Piffer, D. The personality and cognitive correlates of creative achievement. *Open Differential Psychology*. 2014. Retrieved from: Retrieved from <http://openpsych.net/ODP/2014/04/the-personality-and-cognitive-correlates-of-creative-achievement/>
- R Core Team. *R: A language and environment for statistical computing*. Wien, Austria: the R Foundation for Statistical Computing; 2017. Retrieved from: <http://www.R-project.org/>
- Raven, JC. *Advanced Progressive Matrices Set I and II*. H.K. Lewis & Co. , editor. London: 1965.
- Rizopoulos D. 2006; ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*. 17:1–25.
- Runco MA, Okuda SM, Thurston BJ. 1987; The psychometric properties of four systems for scoring divergent thinking tests. *Journal of Psycho-educational Assessment*. 2:149–156.
- Salvi C, Bricolo E, Franconeri SL, Kounios J, Beeman M. 2015; Sudden insight is associated with shutting out visual inputs. *Psychonomic Bulletin & Review*. 22:1814–1819. [PubMed: 26268431]
- Salvi C, Bricolo E, Kounios J, Bowden E, Beeman M. 2016; Insight solutions are correct more often than analytic solutions. *Thinking & Reasoning*. 22:443–460. [PubMed: 27667960]
- Salvi C, Cristofori I, Grafman J, Beeman M. 2016; The politics of insight. *The Quarterly Journal of Experimental Psychology*. 69:1064–1072. [PubMed: 26810954]
- Salvi C, Costantini G, Bricolo E, Perugini M, Beeman M. 2015; Validation of Italian rebus puzzles and compound remote associate problems. *Behavior Research Methods*. 48:664–685.
- Schwarz G. 1978; Estimating the dimension of a model. *The Annals of Statistics*. 6:461–464.
- Schwert PM. 2007; Using sentence and picture clues to solve verbal insight problems. *Creativity Research Journal*. 19:293–306.
- Segal E. 2004; Incubation in insight problem solving. *Creativity Research Journal*. 16(1):141–148.

- Spinnler H, Tognoni G. 1987; Standardizzazione e taratura italiana di test neuropsicologici. *Italian Journal of Neurological Sciences*. 6:54–62.
- Stone CA. 2000; Monte Carlo based null distribution for an alternative goodness-of-fit test statistic in IRT models. *Journal of Educational Measurement*. 37(1):58–75.
- Taft R, Rossiter JR. 1966; The Remote Associates Test: divergent or convergent thinking? *Psychological Report*. 19:1313–1314.
- Vellante M, Zucca G, Preti A, Sisti D, Rocchi MBL, Akiskal KK, et al. 2011; Creativity and affective temperaments in non-clinical professional artists: an empirical psychometric investigation. *Journal of Affective Disorders*. 135:28–36. [PubMed: 21820743]
- Wilcox, R. *Modern statistics for the social and behavioral sciences: a practical introduction*. Boca Raton, FL: CRC press; 2012.
- Wilson, M. *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum; 2005.
- Wright B, Panchapakesan N. 1969; A procedure for sample-free item analysis. *Educational and Psychological Measurement*. 29:23–48.

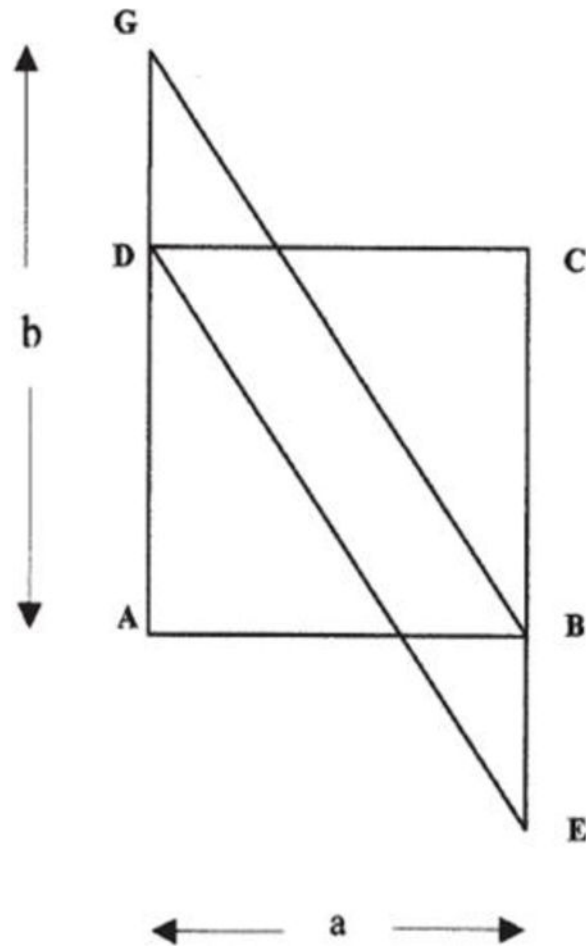


Figure 1.

The “parallelogram” mathematical problem: given that $AB = a$ and $AG = b$, find the sum of the areas of square $ABCD$ and parallelogram $EBGD$. The solution is to restructure the given shape into two partially overlapping triangles: ABG and ECD . The sum of their areas is $2 \times ab/2 = ab$.

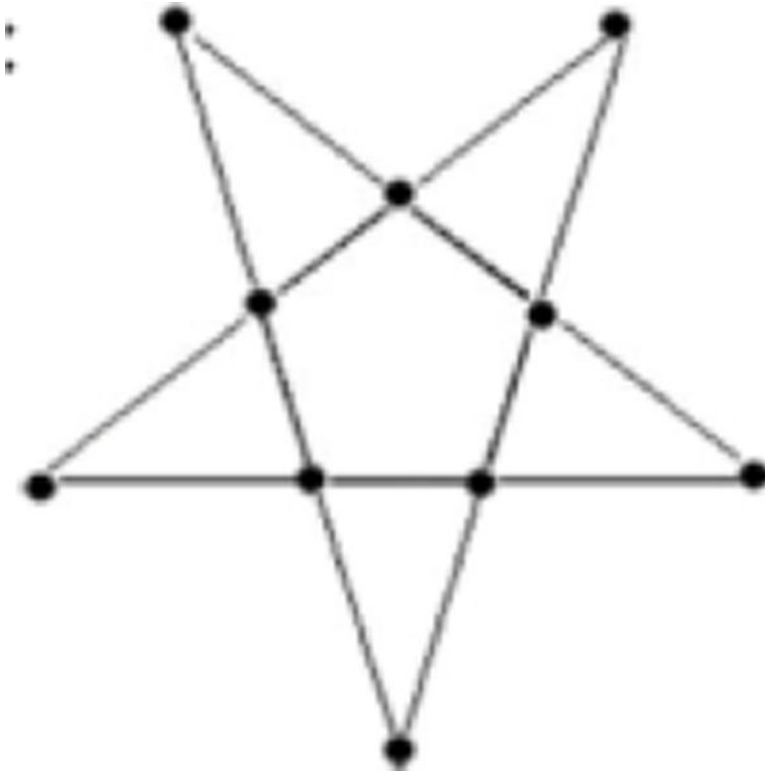


Figure 2. Solution of the “Star” visuo-spatial problem: rows must be arranged in a five-pointed star.

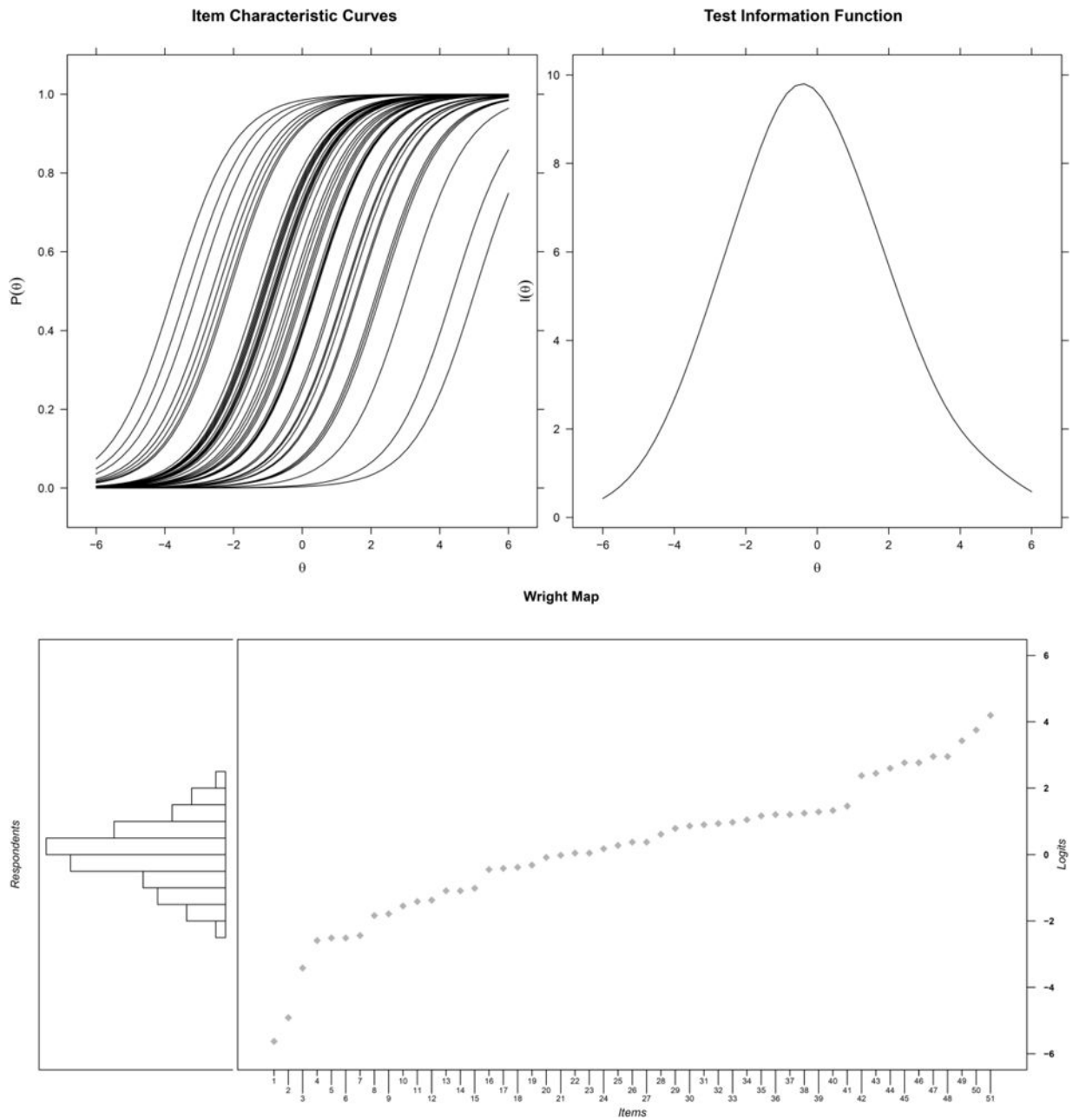


Figure 3. Item characteristic curves, test information function, and Wright map of the final 51-item RAT. For item numbers in the Wright map refer to Table 2.

Table 1

Fit statistics for the 1PL and the 2PL IRT models

Model	Log Likelihood	N. of parameters	AIC	BIC
78-item 1PL	-5304.100	79	10766.20	11005.6
78-item 2PL	-5214.224	156	10740.45	11213.2
51-item 1PL	-3479.062	52	7117.24	7219.70
51-item 2PL	-3446.785	102	7517.81	7406.67

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

IRT parameters for the final 51-item Italian RAT.

#	Item	Solution	Diff.	Infit	z (infit)	Outfit	z (outfit)	χ^2 *	p-value
1	testa.fila.colpo	Coda	-5.63	0.98	0.29	1.23	0.55	1.83	.47
2	sotto.valutazione.errore	Stima	-4.91	1.08	0.34	1.23	0.54	1.60	.57
3	politico.membro.circolo	Socio	-3.42	0.98	0.03	0.54	-0.80	1.77	.63
4	poli.esperto.commissario	Tecnico	-2.59	0.98	-0.05	0.73	-0.66	0.97	.91
5	pidi.blocco.posizione	Stallo	-2.51	0.96	-0.16	0.79	-0.51	3.47	.29
6	via.canale.uomo	Dotto	-2.51	1.02	0.18	0.92	-0.12	1.88	.61
7	mento.miscela.metallo	Lega	-2.44	0.86	-0.75	0.54	-1.51	4.73	.15
8	foto.riassunto.libro	Sintesi	-1.84	0.86	-1.04	0.66	-1.51	4.58	.15
9	visivo.sonoro.canale	Audio	-1.79	1.15	1.12	1.04	0.24	6.26	.07
10	capo.liquidazione.contante	Saldo	-1.55	0.87	-1.17	0.72	-1.42	2.87	.37
11	manica.rotazione.lancette	Giro	-1.41	0.94	-0.50	0.84	-0.84	2.08	.56
12	porta.cibo.arabo	Pane	-1.37	1.04	0.38	1.12	0.66	3.00	.34
13	poli.sanitario.colloquio	Clinico	-1.09	0.95	-0.56	0.99	-0.02	3.43	.27
14	porta.lettura.accesso	Chiave	-1.09	1.02	0.21	1.08	0.56	2.31	.54
15	capo.viso.mascherato	Volto	-1.01	0.92	-0.95	0.78	-1.53	3.51	.27
16	panca.baule.risonanza	Cassa	-0.45	0.96	-0.55	0.88	-1.11	2.36	.48
17	vite.battuta.tesoro	Caccia	-0.42	0.97	-0.39	0.94	-0.54	1.20	.85
18	aroma.trattamento.urto	Terapia	-0.38	1.01	0.11	1.02	0.22	4.08	.20
19	foto.duplicato.testo	Copia	-0.32	0.97	-0.35	1.02	0.21	1.32	.80
20	campo.divino.natale	Santo	-0.08	1.10	1.51	1.12	1.23	4.68	.12
21	super.fiera.coperto	Mercato	-0.02	0.93	-1.06	0.85	-1.53	2.65	.41
22	tra.tribunale.italico	Foro	0.05	1.01	0.19	1.01	0.11	4.17	.18
23	video.convegno.stampa	Conferenza	0.05	0.98	-0.23	0.98	-0.20	3.61	.24
24	toma.conteggio.corrente	Conto	0.18	1.06	0.93	1.03	0.32	4.35	.15
25	via.sedile.elettorale	Seggio	0.28	1.07	0.99	1.05	0.54	3.82	.22
26	capo.mestiere.gruppo	Lavoro	0.38	0.93	-1.04	0.87	-1.30	3.80	.20
27	goccia.conteggio.ore	Re	0.38	0.86	-2.00	0.82	-1.83	2.80	.36
28	auto.pulizia.secco	Lavaggio	0.62	1.02	0.31	1.01	0.13	1.76	.66

#	Item	Solution	Diff.	Infit	z (infit)	Outfit	z (outfit)	χ^2_{2*}	p-value
29	frutta.vivaio.botanico	Orto	0.79	0.95	-0.53	0.86	-1.07	1.57	.75
30	corto.pista.elettrico	Metraggio	0.86	0.91	-1.06	0.90	-0.75	1.53	.76
31	perla.mamma.santa	Madre	0.90	1.05	0.56	1.08	0.63	4.39	.14
32	posta.tavolo.scuola	Banco	0.94	1.04	0.44	0.95	-0.33	2.91	.37
33	vale.alimento.macello	Carne	0.97	0.96	-0.47	0.95	-0.29	4.37	.14
34	bietola.noia.pizzo	Barba	1.05	0.94	-0.60	0.96	-0.21	2.74	.41
35	meno.intervallo.riflessione	Pausa	1.17	0.95	-0.47	0.89	-0.62	2.54	.46
36	sotto.ventre.aria	Pancia	1.21	0.92	-0.80	0.82	-1.12	1.15	.86
37	volo.sfera.neve	Palla	1.21	1.03	0.33	0.92	-0.44	3.34	.29
38	auto.affitto.film	Noleggio	1.25	1.11	1.07	1.14	0.82	4.11	.20
39	calce.uccello.uova	Struzzo	1.29	0.95	-0.50	0.96	-0.15	3.56	.26
40	pesce.lama.guerrero	Spada	1.33	0.88	-1.16	0.78	-1.26	4.58	.14
41	gira.stella.tramonto	Conta	1.46	0.91	-0.76	0.77	-1.20	3.56	.26
42	servo.volante.blocco	Sterzo	2.38	0.99	0.00	0.97	0.03	1.15	.85
43	cottura.spigolo.retto	Angolo	2.45	0.85	-0.82	0.69	-0.90	4.82	.13
44	cavallo.formaggio.pepe	Cacio	2.60	0.87	-0.63	0.50	-1.53	5.74	.10
45	ago.iniezione.insetto	Puntura	2.77	0.98	-0.02	0.74	-0.55	3.67	.25
46	tele.quotidiano.articolo	Giornale	2.77	0.95	-0.16	0.67	-0.79	2.41	.48
47	balla.colpo.rigore	Calcio	2.95	0.96	-0.10	0.66	-0.71	2.07	.54
48	famiglia.abitazione.cura	Casa	2.95	1.00	0.09	1.09	0.35	1.07	.87
49	sotto.gradinata.chiocciola	Scala	3.43	1.01	0.14	1.00	0.19	0.99	.87
50	igienica.foglio.identita	Carta	3.75	1.05	0.27	1.13	0.41	2.71	.38
51	auto.ordigno.atomica	Bomba	4.19	1.07	0.30	0.86	0.06	0.79	.92

Note. Items are presented in order of difficulty (Diff). Since we fitted a IPL model, all items have the same value of the discrimination parameter (1.12). χ^2_{2*} and p-value respectively indicate Stone's (2000) scaled chi square statistic and the corresponding bootstrap-based p-value.

Table 3

Descriptive statistics and correlations.

	M	Mdn	SD	Skew.	Kurt.	α	1	3	5	6	7	8	9	10
1. RAT	0.00	0.03	0.95	-0.05	2.82	.89	1	.50***	.31***	.38***	.11	.53***	.50***	.31***
3. Anagrams	0.00	-0.18	0.96	0.52	3.60	.92	.46***	1	.19*	.16*	.07	.17*	.17*	.18*
5. Raven SPM	45.50	46.00	6.67	-0.89	4.00	.85	.35***	.23**	1	.27***	.06	.15 \dagger	.13	.14 \dagger
6. Classic insight pr.	1.57	2.00	0.70	0.11	2.72	.38	.41***	.19**	.27***	1	.16 \dagger	.30***	.31***	.20*
7. AUT – originality	3.59	2.96	2.63	0.94	3.16	–	.10	.01	.07	.19*	1	.05	.10	.06
8. AUT – flexibility	7.72	7.00	3.55	0.40	2.36	–	.49***	.13	.14 \dagger	.33***	.10	1	.95***	.34***
9. AUT – fluency	9.69	9.00	4.96	0.46	2.59	–	.46***	.14 \dagger	.14 \dagger	.31***	.15 \dagger	.93***	1	.37***
10. CAQ	8.57	6.00	8.35	2.02	8.86	.46	.27***	.08	.15 \dagger	.13	-.03	.33***	.36***	1

Note.

 \dagger $p < .10$,* $p < .05$,** $p < .01$,*** $p < .001$

Pearson correlations are reported below the main diagonal; Spearman correlations are above the diagonal.

Skew. = skewness; Kurt. = kurtosis; α = Cronbach's alpha; RAT = Remote Associates Test; SPM = Standard Progressive Matrices; AUT = Alternative Uses Task; CAQ = Creative Achievement Questionnaire.