The factor structure of the Verbal Torrance Test of Creative Thinking in an Arabic context : classical test theory and multidimensional item response theory analyses

Reference:

Said-Metwaly Sameh, Kyndt Eva, Van den Noortgate Wim.- The factor structure of the Verbal Torrance Test of Creative Thinking in an Arabic context : classical test theory and multidimensional item response theory analyses
Thinking skills and creativity - ISSN 1871-1871 - 35(2020), 100609
Full text (Publisher's DOI): https://doi.org/10.1016/J.TSC.2019.100609
To cite this reference: https://hdl.handle.net/10067/1677280151162165141

The Factor Structure of the Verbal Torrance Test of Creative Thinking in an Arabic Context: Classical Test Theory and Multidimensional Item Response Theory Analyses

Sameh Said-Metwaly, Eva Kyndt, Wim Van den Noortgate

PII: S1871-1871(19)30257-3

DOI: https://doi.org/10.1016/j.tsc.2019.100609

Reference: TSC 100609

To appear in: *Thinking Skills and Creativity*

Received Date: 22 August 2019

Revised Date: 24 October 2019

Accepted Date: 17 November 2019

Please cite this article as: Said-Metwaly S, Kyndt E, Van den Noortgate W, The Factor Structure of the Verbal Torrance Test of Creative Thinking in an Arabic Context: Classical Test Theory and Multidimensional Item Response Theory Analyses, *Thinking Skills and Creativity* (2019), doi: https://doi.org/10.1016/j.tsc.2019.100609

The Factor Structure of the Verbal Torrance Test of Creative Thinking in an Arabic Context: Classical Test Theory and Multidimensional Item Response Theory Analyses

Sameh Said-Metwaly [1,2], Eva Kyndt [1,3], Wim Van den Noortgate [1,2]

[1] KU Leuven (University of Leuven)

[2] imec-itec, KU Leuven (University of Leuven)

[3] University of Antwerp

Correspondence concerning this article should be addressed to Sameh Said-Metwaly, Faculty of Psychology and Educational Sciences, KU Leuven (University of Leuven), 8500 Kortrijk, Belgium. E-mail: sameh.metwaly@kuleuven.be

Research Highlights

- This is the first study of its kind in an Arabic context
- Factor structure of the Arabic Verbal TTCT is assessed in an Egyptian sample
- Classical and multidimensional IRT analyses suggest a six-bifactor structure
- Findings offer valuable implications for creativity testing

Abstract

The Verbal Torrance Test of Creative Thinking (Verbal TTCT) is one of the most frequently used instruments in creativity research. However, its factor structure is still debated and mostly assessed in western contexts. The purpose of the current study was to provide new insights into the factor structure underlying this instrument, with a particular focus on an Arabic context. A sample of 621 undergraduate student from Egypt completed an Arabic version of the Verbal TTCT (Form B). Exploratory and confirmatory factor analyses based on classical test theory were performed, followed by a confirmatory multidimensional item response theory analysis. The findings suggested a novel bifactor structure comprising a general factor in addition to six activity-specific factors. Further support for the bifactor structure was provided by subsequent analyses of residual correlations, empirical reliability, item fit, and item parameters. Theoretical and applied implications of these findings are discussed.

Keywords: Verbal Torrance Test of Creative Thinking, construct validity, factor structure, multidimensional item response theory

The Factor Structure of the Verbal Torrance Test of Creative Thinking in an Arabic Context: Classical Test Theory and Multidimensional Item Response Theory Analyses

1. Introduction

The measurement of creativity is an area of increased research interest as human survival and progress heavily rely upon creative abilities (Hennessey & Amabile, 2010; Puccio, 2017). Over the last decades, the fast-growing body of creativity research has been accompanied by the demand for appropriate approaches to measuring creativity (Kim, Cramond, & Bandalos, 2006). This has led to the application of various approaches, which include divergent thinking tests, personality inventories, and product measures (Clapham, 2004; Said-Metwaly, Kyndt, & Van den Noortgate, 2017a). Among these, divergent thinking tests have largely been used to estimate divergent production, a hypothesized cognitive process underlying creativity (Baer, 2011; Guilford, 1967). Guided by Guilford’s (1956, 1975) structure of intellect model, divergent thinking tests include open-ended stimuli that solicit multiple different responses rather than a unique response. These responses are typically assessed for fluency (generating numerous responses), flexibility (generating responses of various categories), and originality (generating unusual responses; Guilford, 1966).

The Torrance Test of Creative Thinking (TTCT; Torrance, 1966), a measure of divergent thinking, has gained widespread popularity in creativity research. There are two formats of the TTCT, Verbal and Figural, each with two parallel forms, A and B (Torrance & Haensly, 2003). The TTCT has been employed in more than 2,000 published studies (Torrance, 2000) and is available in over 35 languages (Kapoula & Vernet, 2016; Millar, 2002). It can be administered to both young children and adolescents (Torrance & Haensly, 2003).

Despite its long history of use, the validity of the TTCT is an issue of continuing controversy (Said-Metwaly, Fernández-Castilla, Kyndt, & Van den Noortgate, 2018). Much of this controversy is about its construct validity (Bart, Hokanson, & Can, 2017; Clapham, 1998; Zeng, Proctor, & Salvendy, 2011). Questions have been raised as to whether the TTCT really measures the hypothetical underlying creativity factors (Almeida, Prieto, Ferrando, Oliveira, & Ferrándiz, 2008; Clapham, 1998; Krumm, Lemos, & Filippetti, 2014). Yet, there are few studies investigating the latent structure of the TTCT, particularly those focusing on the Verbal format (Krumm, Aranguren, Filippetti, & Lemos, 2014; Said-Metwaly et al., 2018).

Previous studies on the structure of the Verbal TTCT have yielded equivocal findings (Ferrándiz, Ferrando, Soto, Sáinz, & Prieto, 2017; Kim, 2006a, 2006b; Krumm, Filippetti, Lemos, Koval, & Balabanian, 2016). Some researchers have proposed a unidimensional structure for this measure based on the high correlations found between its subscale scores (e.g., Dixon, 1979; Hocevar, 1979; Hocevar & Michael, 1979). Other researchers have proposed a multidimensional structure. By applying exploratory factor analysis (EFA) to Verbal and Figural

TTCT data, Plass, Michael, and Michael (1974) identified seven factors representing the particular demand of each activity and not the hypothetical creativity construct. Moreover, through principal component analysis (PCA), Clapham (2004) concluded that the subscale scores of the Verbal and Figural TTCT reflected two independent factors associated with the format of the activities. In line with these findings, using PCA in three studies carried out in Spain and Portugal, Almeida et al. (2008) distinguished multiple factors (three, five, and six factors) related to the demand and format of the activities. Krumm, Aranguren, Filippetti, and Lemos (2014) also analyzed the Verbal TTCT data using confirmatory factor analysis (CFA) and found six factors related to the demand of each activity. In sum, the existing literature provides conflicting evidence regarding the latent structure of the Verbal TTCT.

2. The Present Study

As outlined earlier, there are scanty and contradictory research findings regarding the construct validity of the Verbal TTCT. In addition, the construct validity of this instrument has been mostly assessed in western contexts (Said-Metwaly, Kyndt, & Van den Noortgate, 2017b). The lack of published validation studies of this instrument in different cultural contexts seems to be a significant research gap. It remains unclear so far what validity the western measures of psychological constructs like creativity have when exported to other cultural contexts (Humble, Dixon, & Mpofu, 2018; Mpofu, Myambo, Mogaji, Mashego, & Khaleefa, 2006). Besides, as far as we know, there are no validation studies of this instrument using item response theory methods. This is regrettable as such modern psychometric analyses could offer additional insights into instrument and item functioning (Plucker & Makel, 2010; Said-Metwaly, Kyndt, & Van den Noortgate, 2017b). To that end, this study sought to validate the dimensionality of the Verbal TTCT in an Arabic context using both classical test theory (CTT) and multidimensional item response theory (MIRT) analyses.

3. Method

3.1. Subjects

Subjects were volunteer undergraduate students from Damanhour University in Egypt. They were enrolled in educational psychology courses at the Faculty of Education. All of them were Arabic native speakers. The original sample included 649 subjects (515 females). The gender imbalance in the sample reflects the gender imbalance in the target population. Twenty-eight subjects were excluded for providing incomplete responses, leaving a total of 621 subjects (484 females). Ages ranged from 19 to 22, with a mean age of 19.77 (*SD* = 2.33). The principals of Damanhour University provided permission for this data collection.

3.2. Measure

The Verbal Torrance Test of Creative Thinking (TTCT; Torrance, 1966)

The Arabic translated version of the Verbal TTCT (Form B) was administered (Soleimon & Abu Hatab, 1973). There were six activities with varying time limits: asking (writing down questions about an ambiguous pictured event, 5 min); guessing causes (listing possible causes underlying the event, 5 min); guessing consequences (postulating possible results of the event, 5 min); product improvement (suggesting improvements to a toy monkey, 10 min); unusual uses (proposing alternative uses for tin cans, 10 min); and just suppose (listing consequences for a certain improbable situation, 5 min; Torrance, 1967; Torrance & Haensly, 2003). The original version of Verbal TTCT involved one more activity (the unusual questions activity), but this activity was eliminated in the current version as it did not add further meaningful information on an individual's divergent thinking (Cramond, Matthews-Morgan, Bandalos, & Zuo, 2005; Kaufman, Plucker, & Russell, 2012; Kim, 2017).

3.3. Data collection and scoring

Data collection took place in classrooms at Damanhour University over a period of about 3 weeks. The Verbal TTCT was administered in a group setting with approximately 50 students. Testing was conducted by five trained research assistants under the supervision of the first author. It was made clear to all subjects that participation was voluntary, and that the data collected would remain confidential. Subjects were informed that, by participation in the study, they could gain knowledge about how creativity is measured and get their creativity scores. After providing written informed consent, subjects completed the test. Subjects were asked to give as many responses as possible to each of the test activities within the specified time limit. They had the opportunity to ask clarification questions to ensure a clear understanding of the demand of each activity. They were allowed to leave after completing the whole test.

After excluding irrelevant, unclear, and identical responses, the subject's responses to each activity were scored for fluency, flexibility, and originality. Fluency represented an estimate of how many responses were generated. Flexibility represented an estimate of how many distinct categories were reflected in responses. Originality was determined depending on the percentage of subjects who generated each response; two points were given for responses with a frequency percentage lower than 2%; one for those with a percentage between 2-5%; and zero for those with a percentage greater than 5%.

3.4. Analyses

Given that previous validation studies of the Verbal TTCT have used EFA and CFA based on CTT, similar analyses were employed in the current study to allow for comparability of results. In addition, MIRT analyses were applied to explore whether the identified factor structure can be replicated under an alternative statistical framework, and also to allow for further psychometric evaluation of instrument and item functioning.

MIRT models are item response theory models that assume more than one dimension underlying responses to a particular set of items, therefore modeling each item in a continuous multidimensional space (Kacmar, Farmer, Zivnuska, & Witt, 2006). MIRT and CTT analyses differ in many aspects. First, the relationship between indicator and latent variables in CTT analyses is assumed to be linear, and consequently, one number (i.e., factor loading) is used to represent this relationship along all latent variable levels. On the contrary, this relationship in MIRT analyses could be nonlinear, and therefore information on this relationship are provided across all latent variable levels (Greguras, 2005). Second, MIRT overcomes the item-person confounding problem faced by CTT (Osteen, 2010a, 2010b). While person and item parameter estimates are sample-dependent in CTT analyses, MIRT analyses yield sample-free estimates. As a result, MIRT analyses seek to find the ultimate solution that adapts the same latent space across various samples and tests (Reckase, 2009). Third, while the standard error of measurement is constant and sample-dependent in CTT analyses, it is variant across latent trait levels and population-general in MIRT analyses (Osteen, 2010a). Finally, in addition to model fit indices produced from CTT analyses, MIRT analyses provide estimates of item fit, item parameters, and item and instrument information functions (Osteen, 2010a).

The total sample was initially split at random into two subsamples. The first subsample ($n = 310$) was used to conduct exploratory analyses, and the second subsample ($n = 311$) to conduct confirmatory analyses. In this way, the instrument’s factor structure can be explored in one subsample and tested in a separate subsample. In situations when sample size permits, this procedure is recommended in order to cross-validate the findings and ensure their stability across independent subsamples (Brown, 2015; Byrne, 2010; Everett, 1983; Fabrigar, Wegener, MacCallum, & Strahan, 1999). In addition, implementing exploratory and confirmatory analyses with randomly split subsamples enables taking advantage of the differences in emphasis between these analyses (Fabrigar & Wegener, 2014).

In the exploratory step, the appropriateness of the data for structure detection was first tested by Bartlett's test of sphericity and Kaiser-Meyer-Olkin measure of sampling adequacy. A significant Bartlett's test (Bartlett, 1954) and a value of Kaiser-Meyer-Olkin measure exceeding .60 (Kaiser, 1974) are required. EFA using principal axis factoring with oblimin rotation was then conducted. Identification of optimal number of factors was based upon Kaiser's (1960) eigenvalue larger than one criterion, Cattell's (1966) scree plot, and Horn's (1965) parallel analysis. The optimal coordinates and acceleration factor methods (Raiche, Walls, Magis, Riopel, & Blais, 2013) were also used to obtain numerical solutions to the scree plot. The acceleration factor identifies the most abrupt shift (known as "elbow") in the slope of the scree plot, keeping the factors preceding this shift. The optimal coordinates method uses linear regression models connecting the smallest eigenvalue to each of the remaining eigenvalues, retaining all factors with eigenvalues greater than the predicted eigenvalues. The EFA was carried out with the *nFactors* (Raiche & Magis, 2015) and *psych* (Revelle, 2017) packages in R.

For the classical CFA, maximum likelihood estimation was used with the *lavaan* package (Rosseel, 2012). Five hypothesized models of the Verbal TTCT were tested (see Figure 1). The first model proposed a unidimensional factor structure. The second model proposed a three correlated factor structure comprising fluency, flexibility, and originality, in which the corresponding indicators across activities loaded on the same factor. The third model proposed a six correlated factor structure in which the three indicators of each activity loaded on a separate factor. The two remaining models proposed bifactor structures. Evaluation of bifactor models is a recommended practice when exploring construct dimensionality (Reise, Morizot, & Hays, 2007). The first bifactor model had a general factor and three specific factors (fluency, flexibility, and originality). The second bifactor model had a general factor and six specific factors representing the activities. The suggested bifactor models correspond to the hybrid conceptualization of creativity, which assumes that creativity includes both general and domain-specific abilities (Amabile, 1983; Baer & Kaufman, 2005; Plucker & Beghetto, 2004; Sternberg, 1989). According to this conceptualization, an individual may have a general creative ability that contributes to performance on different tasks. An individual may also have specific creative abilities that apply to particular tasks. Following this conceptualization, one can assume that the variance of each indicator of the Verbal TTCT can be partitioned into that common to all indicators and that unique to a cluster of indicators related in some way (i.e., similar category or activity).

A confirmatory MIRT analysis was subsequently carried out with the *mirt* package (Chalmers, 2012). Samejima's (1969) graded response model was used as an appropriate model for the analysis of polytomous data. In this model, each item is characterized by a single discrimination (*a*) parameter along with multiple difficulty (b_1, b_2, b_3, etc.) parameters based on the item score range. Considering the high-dimensional structure of the Verbal TTCT, Cai's (2010) Metropolis-Hastings Robbins-Monro algorithm was used to obtain parameter estimates.

An acceptable model fit was determined based on smaller and insignificant chi-square (χ^2), normed chi-square (χ^2/df) below 5, comparative fit index (CFI) exceeding .95, Tucker Lewis index (TLI) exceeding .95, root mean square error of approximation (RMSEA) below .06, and standardized root mean square residual (SRMR) below .08 (Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1999). In the MIRT analyses, Cai and Monroe's (2014) C_2 statistic was computed as a substitute for χ^2. As a limited-information statistic, C_2 is well calibrated and could substantially outperform full-information statistics (like χ^2) in identifying model misspecification (Cai & Monroe, 2014). Smaller and non-significant values of C_2 are preferable. To aid in the evaluation of model fit, the following information criteria were also reported: Akaike information criterion (AIC), Bayesian information criterion (BIC), and sample-size adjusted BIC (SABIC). Lower values of these information criteria indicate a better fit (Burnham & Anderson, 2003).

The best-fitting model was subjected to a detailed MIRT analysis to evaluate residual correlations, empirical reliability, item fit, and item parameters. Absolute values of residual correlation up to .20 with an absolute average below .10 are considered acceptable (Amtmann et al., 2010). Item fit was assessed using the signed chi-squared test (S-χ^2; Orlando & Thissen, 2000, 2003), with significant values indicating misfit. Since MIRT models rarely fit perfectly (Steinberg & Thissen, 2013; Zhang & Stone, 2008) and also large samples tend to yield significant χ^2 (Stone & Zhang, 2003), item misfit was identified using the .01 level of significance rather than the .05 level. An item commonly provides a greater amount of information on latent ability when it has a wider range of difficulty parameters (Huang et al., 2017). Discrimination values below 0.34 indicate very low discrimination; values between 0.35 and 0.64 indicate low discrimination; values between 0.65 and 1.34 indicate moderate discrimination; values between 1.35 and 1.69 indicate high discrimination; and values above 1.69 indicate very high discrimination (Baker, 2001).

4. Results

4.1. CTT analyses

4.1.1. EFA results

Bartlett's test ($\chi^2 = 4122.11$, $df = 153$, $p < .001$) and Kaiser-Meyer-Olkin measure (.78) indicated that the data were factorable. As shown in Figure 2, Kaiser's criterion, parallel analysis, and optimal coordinates yielded a six-factor solution, while the acceleration factor method yielded a single-factor solution. The one factor solution did gain further support from the significant positive correlations observed between indicators across activities as well as between indicators and the total score (see Supplementary Table S1). EFA was thus conducted, fitting one- and six-factor solutions. Table 1 reports the results of the classical EFA. The one-factor solution explained 29.88% of the total variance, with adequate standardized factor loadings ranging from .31 to .70. However, the fit indices obtained for this model were unsatisfactory (see Table 1). The six-factor solution explained 70.27% of the variance. Each of the six factors grouped three indicators of a single activity, and correlations between factors ranged from .21 to .44. While the six-factor solution better fitted the data, it produced four Heywood cases (i.e., factor loadings above 1.00) that often suggest model misspecification (Brown, 2015). Given this, an orthogonal varimax rotation of the six-factor solution was also conducted for comparison, yet the results did not substantially differ and the Heywood cases remained. As such, the results of the orthogonal rotation were not reported.

The possibility of a three-bifactor or six-bifactor solution was tested using Jennrich and Bentler's (2011) exploratory bifactor rotation in the *psych* package. The three-bifactor solution explained 56.89% of the total variance. The indicators of activity 6 were grouped in specific factor 1; the indicators of activities 2, 4, and 5 were in specific factor 2; and the indicators of activities 1 and 3 were in specific factor 3. Absolute factor loadings ranged from .33 to .71 for the general factor and from .26 to .83 for the specific factors. The indicators of activities 1 and 2 showed negative factor loadings. The fit indices for this model were unsatisfactory. The six-bifactor solution could explain 74.56% of the variance. The grouping of indicators in this solution was identical to the hypothesized six-bifactor model. Apart from the loading of the indicator (Ori3) on its specific factor (.24), all loadings were adequate and ranged from .33 to .69 for the general factor and from .40 to .89 for the specific factors. The six-bifactor solution showed superior fit indices, which supported the presence of common and unique variance among the indicators.

4.1.2. CFA results

Table 2 reports the results of the classical CFA. The three-bifactor model failed to converge; therefore, the results for this model were not reported. The one-factor model did not fit the data well nor did the three-factor model, although yielding acceptable factor loadings. Additionally, the three-factor model was troubled by Heywood cases as all factor correlations were greater than 1.00. The six-factor and six-bifactor models revealed similarly acceptable fit. However, five factor loadings in the six-factor model exceeded 1.00, again representing Heywood cases. Thus, the six-factor model was not statistically admissible. On the contrary, all factor loadings for the six-bifactor model fell within the acceptable range (from .35 to .66 for the general factor and from .37 to .92 for the specific factors).

4.2. MIRT Results

Table 3 reports the results of the MIRT analysis. The one-factor, three-factor, and three-bifactor models demonstrated a poor fit. Interestingly, while the six-factor model had an improved fit, it failed to satisfy the minimum fit requirements. It is worth noticing here that no Heywood cases were observed for the three-factor or six-factor models, yet in some cases, the values of loadings or factor correlations approached 1.00. The six-bifactor model exhibited the best fit to the data, as well as acceptable factor loadings ranging from .35 to .66 for the general factor and from .37 to .86 for the specific factors.

Based on all results, the six-bifactor model seemed to provide the best representation of the Verbal TTCT structure among the tested models and was, therefore, subjected to further MIRT analyses. In this model, all indicators showed adequate residual correlations less than the absolute value of .20, with an average residual correlation of .02. The residual correlation matrix is available upon request from the corresponding author. Empirical reliability for this model was found to be high for both the general factor (.90) and the six specific factors (.89, .89, .90, .92, .87, .93, respectively). Item fit and item parameter estimates for the six-bifactor model are reported in Supplementary Table S2. All indicators demonstrated an acceptable fit as none of the S-χ^2 values was associated with p below .01 or RMSEA above .06. Due to differences in score range, the indicators had dissimilar numbers of difficulty estimates. The difficulty estimates ranged from -47.61 to 37.22 for fluency indicators, -16.90 to 14.72 for flexibility indicators, and -7.75 to 0.46 for originality indicators. This suggests that fluency and flexibility indicators provide information on a wider range of ability levels than originality indicators. The discrimination estimates were very high (range, 2.90 to 15.62) for fluency indicators, high to very high (range, 1.58 to 5.37) for flexibility indicators, and moderate to high (range, 0.70 to 1.45) for originality indicators.

5. Discussion

The aim of the current study was to explore the latent factor structure of the Verbal TTCT in an Arabic-speaking sample using CTT and MIRT analyses. Alternative models of the Verbal TTCT were tested. The previously suggested one-factor (Dixon, 1979; Hocevar, 1979; Hocevar & Michael, 1979) and three-factor (Torrance, 1966) models failed to provide an appropriate fit to the data. Although the six-factor model (Krumm, Aranguren, Filippetti, & Lemos, 2014) offered an improved fit, it had problems with Heywood cases and was also not supported by the MIRT analyses. It is important to note that factor loadings equaling or approaching 1.00 were also reported by Krumm, Aranguren, Filippetti, and Lemos (2014) for the six-factor model; however, they concluded that their data supported that model. In the current study, the best fit across all analyses was achieved by a proposed bifactor model having a general factor as well as six distinct factors representing the activities of the instrument. This was also evident from the acceptable residual correlations, empirical reliability, item fit, and item parameters obtained for this model.

The current study adds to the growing body of knowledge on the measurement of creativity in many ways. First, compared to previous validation studies primarily using western subjects, the current study was conducted with Arabic-speaking subjects and can therefore add new insights into the construct validity of the Verbal TTCT. Second, the study sample was sufficiently large and allowed applying EFA and CFA techniques in split-half samples. Importantly, this study also applies, for the first time, MIRT to the Verbal TTCT data. These multiple analytic techniques endorse the robustness of the study findings. Third, the six-bifactor structure discovered in this study suggests that the indicators of the Verbal TTCT reflect some degree of specificity above the common construct represented by the general factor. In such a way, evaluations using this instrument should be based not only on a composite score, but also on scores from individual activities. Along the same lines, it could be argued that the activities of the Verbal TTCT are not fully interchangeable and that utilizing different activities is probably a reason for mixed findings in creativity research. Fourth, this study speaks to the controversial issue of domain-generality/specificity of creativity. In accordance with the hybrid models of creativity, the current findings suggest that general and specific components work together to contribute to creative performance. Finally, the findings presented herein indicated that fluency and flexibility indicators provided a higher discrimination and a wider range of difficulty than did originality indicators. Fluency and flexibility indicators might thus be more appropriate for shortened versions of the Verbal TTCT, because they can effectively differentiate subjects at various levels of the latent construct. Such information might be of a considerable importance to the users of this instrument in research and practice.

The findings of this study are subject to certain limitations that highlight directions for future work. One limitation is that the study’s subjects were all undergraduate students and predominantly female, which might constrain the generalizability of the findings. Further research is thus required to test the current findings in more diverse samples. In the same vein, an issue not addressed in this study was the factorial invariance of the Verbal TTCT across different groups. The validity of the six-bifactor structure proposed here might vary as a function of culture, gender, grade level, or scoring method; therefore, this could be a worthwhile direction for future research. Besides, previous research suggested that environmental factors, such as testing conditions, could affect subject performance as well as psychometric properties of creativity instruments (e.g., Benedek, Mühlmann, Jauk, & Neubauer, 2013; Chand & Runco, 1993; Forthmann, Lips, Szardenings, Scharfen, & Holling, 2018; Said-Metwaly, Fernández-Castilla, Kyndt, & Van den Noortgate, 2019). Hence, even with a controlled data collection process, the effect of external factors, such as instructions and time limits, on the current findings cannot be completely ruled out. In addition to the aforementioned limitations, the Verbal TTCT scoring requires considerable time, effort, and experience (Bart et al., 2017; Clapham, 2004), which is also likely to have affected the resulting data.

In conclusion, this study contributes to the literature on the factor structure underlying the Verbal TTCT, with an emphasis on an Arabic-speaking context. The findings indicate a bifactor structure consisting of a general factor and six activity-specific factors. The suggested bifactor structure has valuable implications for creativity testing and can also serve as an impetus for further work in this research area.

Author Note

Sameh Said-Metwaly, Faculty of Psychology and Educational Sciences & imec-itec, KU Leuven (University of Leuven); Eva Kyndt, Faculty of Psychology and Educational Sciences, KU Leuven (University of Leuven) & Faculty of Social Sciences, University of Antwerp; Wim Van den Noortgate, Faculty of Psychology and Educational Sciences & imec-itec, KU Leuven (University of Leuven).

This research was supported by a grant from the Ministry of Higher Education and Scientific Research in Egypt.

Declarations of interest

None.

Acknowledgements

The authors gratefully acknowledge the contributions of Asmaa Abdelstar, Mohammed M. Elgali, Reda Elshahaat, Mohammed Sami, Nancy Mahmoud, Rehab Sami, and Yasmin Hassan in assisting with data collection in this study.

References

Almeida, L. S., Prieto, M. D., Ferrando, M., Oliveira, E., & Ferrándiz, C. (2008). Torrance Test of Creative Thinking: The question of its construct validity. *Thinking Skills and Creativity, 3*, 53-58. doi: 10.1016/ j.tsc.2008.03.003

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology, 45*, 357-376. doi: 10.1037/0022-3514.45.2.357

Amtmann, D., Cook, K. F., Jensen, M. P., Chen, W.-H., Choi, S., Revicki, D., … Lai, J.-S. (2010). Development of a PROMIS item bank to measure pain interference. *Pain, 150*, 173-182. doi: 10.1016/j.pain.2010.04.025

Baer, J. (2011). How divergent thinking tests mislead us: Are the Torrance Tests still relevant in the 21st century? The division 10 debate. *Psychology of Aesthetics, Creativity, and the Arts, 5*, 309-313. doi: 10.1037/a0025210

Baer, J., & Kaufman, J. C. (2005). Bridging generality and specificity: The Amusement Park Theoretical (APT) model of creativity. *Roeper Review, 27*, 158-163. doi: 10.1080/02783190509554310

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation. Retrieved from https://eric.ed.gov/?id=ED458219

Bart, W. M., Hokanson, B., & Can, I. (2017). An investigation of the factor structure of the Torrance Tests of Creative Thinking. *Educational Sciences: Theory & Practice*, *17*, 515-528. doi: 10.12738/estp.2017.2.0051

Bartlett, M. S. (1954). A note on multiplying factors for various chi square approximations. *Journal of Royal Statistical Society, 16*, 296-298.

Benedek, M., Mühlmann, C., Jauk, E., & Neubauer, A. C. (2013). Assessment of divergent thinking by means of the subjective top-scoring method: Effects of the number of top-ideas and time-on-task on reliability and validity. *Psychology of Aesthetics, Creativity, and the Arts, 7*, 341-349. doi: 10.1037/a0033644

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: The Guildford Press.

Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). New York, NY: Routledge.

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika, 75*, 33-57. doi: 10.1007/s11336-009-9136-x

Cai, L., & Monroe, S. (2014). *A new statistic for evaluating item response theory models for ordinal data* (CRESST Report 839). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276. doi: 10.1207/s15327906mbr0102_10

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*, 1-29. doi: 10.18637/jss.v048.i06

Chand, I., & Runco, M. A. (1993). Problem finding skills as components in the creative process. *Personality and Individual Differences, 14*, 155-162. doi: 10.1016/0191-8869(93)90185-6

Clapham, M. M. (1998). Structure of the Figural Forms A and B of the Torrance Tests of Creative Thinking. *Educational and Psychological Measurement, 58*, 275-283. doi: 10.1177/0013164498058002010

Clapham, M. M. (2004). The convergent validity of the Torrance Tests of Creative Thinking and creativity interest inventories. *Educational and Psychological Measurement, 64*, 828-841. doi: 10.1177/0013164404263883

Cramond, B., Matthews-Morgan, J., Bandalos, D., & Zuo, L. (2005). A report on the 40-year follow-up of the Torrance Tests of Creative Thinking: Alive and well in the new millennium. *Gifted Child Quarterly, 49*, 283-291.

Dixon, J. (1979). Quality versus quantity: The need to control for the fluency factor in originality scores from the Torrance tests. *Journal for the Education of the Gifted, 2*, 70-79.

Everett, J. E. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research, 18*, 197-218. doi: 10.1207/s15327906mbr1802_5

Fabrigar, L. R., & Wegener, D. T. (2014). Exploring causal and noncausal hypotheses in nonexperimental data. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 504-533). New York, NY: Cambridge University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299. doi: 10.1037/1082-989X.4.3.272

Ferrándiz, C., Ferrando, M., Soto, G., Sáinz, M., & Prieto, M. D. (2017). Divergent thinking and its dimensions: What we talk about and what we evaluate? *Annals of Psychology, 33*, 40-47. doi: 10.6018/analesps.32.3.224371

Forthmann, B., Lips, C., Szardenings, C., Scharfen, J., & Holling, H. (2018). Are speedy brains needed when divergent thinking is speeded-or unspeeded? *The Journal of Creative Behavior*. Advance online publication. doi: 10.1002/jocb.350

Greguras, G. J. (2005). Managerial experience and the measurement equivalence of performance ratings. *Journal of Business and Psychology, 19*, 383-397. doi: 10.1007/s10869-004-2234-y

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin, 53*, 267-293. doi: 10.1037/h0040755

Guilford, J. P. (1966). Measurement and creativity. *Theory into Practice, 5*, 186-202.

Guilford, J. P. (1967). *The nature of human intelligence*. New York, NY: McGraw-Hill.

Guilford, J. P. (1975). Creativity: A quarter century of progress. In I. A. Taylor, & J. W. Getzels (Eds.), *Perspectives in creativity* (pp. 37-59). Chicago, IL: Aldine.

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology, 61*, 569-598. doi: 10.1146/annurev.psych.093008.100416

Hocevar, D. (1979). The unidimensional nature of creative thinking in fifth grade children. *Child Study Journal, 9*, 273-278.

Hocevar, D., & Michael, W. B. (1979). The effects of scoring formulas on the discriminant validity of tests of divergent thinking. *Educational and Psychological Measurement, 39*, 917-921. doi: 10.1177/001316447903900427.

Hooper, D., Coughlan, J., Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Business Research Methods, 6*, 53-60.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185. doi: 10.1007/BF02289447

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural *Equation Modeling, 6*, 1-55. doi: 10.1080/10705519909540118

Huang, Y-H-, Lee, J., Chen, Z., Perry, M, Cheung, J. H., & Wang M. (2017). An item-response theory approach to safety climate measurement: The liberty mutual safety climate short scales. *Accident Analysis and Prevention, 103*, 96-104. doi: 10.1016/j.aap.2017.03.015

Humble, S., Dixon, P., & Mpofu, E. (2018). Factor structure of the Torrance Tests of Creative Thinking Figural Form A in Kiswahili speaking children: Multidimensionality and influences on creative behavior. *Thinking Skills and Creativity, 27*, 33-44. doi: 10.1016/j.tsc.2017.11.005

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika, 76*, 537-549. doi:10.1007/s11336-011-9218-4

Kacmar, K. M., Farmer, W. L., Zivnuska, S., & Witt, L. A., (2006). Applying multidimensional item response theory analysis to a measure of meta-perspective performance. *The Electronic Journal of Business Research Methods, 4*, 23-30.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151. doi: 10.1177/001316446002000116

Kaiser, H. F.(1974). An index of factorial simplicity. *Psychometrika, 39*, 13-36. doi: 10.1007/BF02291575

Kapoula, Z., & Vernet, M. (2016). Dyslexia, education and creativity, a cross-cultural study. In Z. Kapoula, & M. Vernet (Eds.), *Aesthetics and neuroscience* (pp. 31-44). New York, NY: Springer.

Kaufman, J. C., Plucker, J. A., & Russell, C. M. (2012). Identifying and assessing creativity as a component of giftedness. *Journal of Psychoeducational Assessment, 30*, 60-73. doi: 10.1177/0734282911428196

Kim, K. H. (2006a). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal, 18*, 3-14. doi: 10.1207/s15326934crj1801_2

Kim, K. H. (2006b). Is creativity unidimensional or multidimensional? Analyses of the Torrance Tests of Creative Thinking. *Creativity Research Journal, 18*, 251-259. doi: 10.1207/s15326934crj1803_2

Kim, K. H. (2017). The Torrance Tests of Creative Thinking - Figural or Verbal: Which one should we use? Creativity. *Theories - Research - Applications, 4*, 302-321. doi: 10.1515/ctra-2017-0015

Kim, K. H., Cramond, B., & Bandalos, D. L. (2006). The latent structure and measurement invariance of scores on the Torrance Tests of Creative Thinking - Figural. *Educational and Psychological Measurement, 66*, 459-477. doi: 10.1177/0013164405282456

Krumm, G., Aranguren, M., Filippetti, V. A., & Lemos, V. (2014). Factor structure of the Torrance Tests of Creative Thinking Verbal Form B in a Spanish-speaking population. *The Journal of Creative Behavior, 50*, 150-14. doi: 10.1002/jocb.76

Krumm, G., Filippetti, V. A., Lemos, V., Koval, J., & Balabanian, C. (2016). Construct validity and factorial invariance across sex of the Torrance Test of Creative Thinking - Figural Form A in Spanish-speaking children. *Thinking Skills and Creativity, 22*, 180-189. doi: 10.1016/j.tsc.2016.10.003

Krumm, G., Lemos, V., & Filippetti, V. A. (2014). Factor structure of the Torrance Tests of Creative Thinking Figural Form B in Spanish-speaking children: Measurement invariance across gender. *Creativity Research Journal, 26*, 72-81. doi: 10.1080/10400419.2013.843908

Millar, G. W. (2002). *The Torrance kids at mid-life*. Westport, CT: Ablex.

Mpofu, E., Myambo, K., Mogaji, A. A., Mashego, T.-A., & Khaleefa, O. H. (2006). African perspectives on creativity. In J. C. Kaufman, & R. J. Sternberg (Eds.), *The international handbook of creativity* (pp. 456-489). New York, NY: Cambridge University Press.

Orlando, M. & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24*, 50-64. doi: 10.1177/01466216000241003

Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2 : An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement, 27*, 289-298. doi: 10.1177/0146621603027004004

Osteen, P. J. (2010a). An introduction to using multidimensional item response theory to assess latent factor structures. *Journal of the Society for Social Work and Research, 1*, 66-82. doi: 10.5243/jsswr.2010.6

Osteen, P. J. (2010b, January). *Multidimensional item response theory analyses as an alternative/supplement to confirmatory factor analyses in evaluating factor structures and measurement invariance*. Paper presented at the 14th Annual Conference of the Society for Social Work and Research, San Francisco, CA.

Plass, H., Michael, J. J., & Michael, W. B. (1974). The factorial validity of the Torrance Tests of Creative Thinking for a sample of 111 sixth-grade children. *Educational and Psychological Measurement, 34*, 413-414.

Plucker, J. A., & Beghetto, R. A. (2004). Why creativity is domain general, why it looks domain specific, and why the distinction does not matter. In R. J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity: From potential to realization* (pp. 153-167). Washington, DC: American Psychological Association.

Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. Kaufman, & R. J.

Sternberg (Eds.), *The Cambridge handbook of creativity*. New York, NY: Cambridge University Press.

Puccio, G. J. (2017). From the dawn of humanity to the 21st century: Creativity as an enduring survival skill. *The Journal of Creative Behavior, 51*, 330-334. doi: 10.1002/jocb.203

Raiche, G., & Magis, D. (2015). *nFactors: Parallel analysis and non graphical solutions to the Cattell Scree Test*. Montreal, Canada: Université du Quebec. Retrieved from https://CRAN.R-project.org/package=nFactors

Raiche, G., Walls, T. A., Magis, D., Riopel, M., & Blais, J.-G. (2013). Non graphical solutions for the Cattell's scree test. *Methodology, 9*, 23-29.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19-31. doi: 10.1007/s11136-007-9183-7

Revelle, W. (2017). *psych: Procedures for psychological, psychometric, and personality research*. Evanston, IL: Northwestern University. Retrieved from https://CRAN.Rproject.org/package=psych

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1-36. doi: 10.18637/jss.v048.i02

Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., & Van den Noortgate, W. (2018). The factor structure of the Figural Torrance Tests of Creative Thinking: A meta-confirmatory factor analysis. *Creativity Research Journal, 30*, 352-360. doi: 10.1080/10400419.2018.1530534

Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., & Van den Noortgate, W. (2019). Testing conditions and creative performance: Meta-analyses of the impact of time limits and instructions. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. doi: 10.1037/aca0000244

Said-Metwaly, S., Kyndt, E., & Van den Noortgate, W. (2017a). Approaches to measuring creativity: A systematic literature review. *Creativity. Theories - Research - Applications, 4*, 238-275. doi: 10.1515/ctra-2017-0013

Said-Metwaly, S., Kyndt, E., & Van den Noortgate, W. (2017b). Methodological issues in measuring creativity: A systematic literature review. *Creativity. Theories - Research - Applications, 4*, 276-301. doi: 10.1515/ctra-2017-0014

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34*, 100-114. doi: 10.1007/BF03372160

Soleimon, A., & Abu Hatab, F. (1973). *Torrance test of creative thinking*. Cairo, Egypt: Egyptian Anglo Library (in Arabic).

Steinberg, L., & Thissen, D. (2013). Item response theory. In J. S. Comer, & P. C. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 336-373). New York, NY: Oxford University Press.

Sternberg, R. J. (1989). Domain-generality versus domain-specificity: The life and impending death of a false dichotomy. *Merrill-Palmer Quarterly, 35*, 115-130.

Stone, C. A., & Zhang, B. (2003). Assessing goodness of fit of item response theory models: A comparison of traditional and alternative procedures. *Journal of Educational Measurement*, *40*, 331-352. doi: 10.1111/j.1745-3984.2003.tb01150.x

Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking: Norms-technical manual*. Lexington, MA: Personal Press.

Torrance, E. P. (1967). *Understanding the fourth grade slump in creative thinking* (Report No. BR-5-0508; CRP-994). Washington, DC: U. S. Office of Education.

Torrance, E. P. (2000). Preschool creativity. In B. A. Bracken (Ed.), *Psychoeducational assessment of preschool children* (3rd ed., pp. 349-363). Needham Heights, MA: Allyn and Bacon.

Torrance, E. P., & Haensly, P. A. (2003). Assessment of creativity in children and adolescents. In Reynoolds, C. R., & Kamphaus, R. W. (Eds.) *Handbook of psychological and educational assessment of children: Intelligence, aptitude and achievement* (2nd ed., pp. 584-607). New York, NY: The Guildford Press.

Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creativity Research Journal, 23*, 24-37. doi: 10400419.2011.545713

Zhang, B., & Stone, C. A. (2008). Evaluating item fit for multidimensional item response models. *Educational and Psychological Measurement, 68*, 181-196. doi: 10.1177/0013164407301547

Table 1

Results of the Classical EFA

Indicators	One-factor solution		Six-factor solution						Three-bifactor solution				Six-bifactor solution						
			F1	F2	F3	F4	F5	F6	G	F1	F2	F3	G	F1	F2	F3	F4	F5	F6
Flu1	.62						1.05		.71			-.55	.57				.89		
Fle1	.56						.76		.62			-.43	.52				.62		
Ori1	.34						.56		.40			-.34	.33				.45		
Flu2	.60				1.01				.69		-.51		.61						.87
Fle2	.57				.83				.64		-.49		.57						.58
Ori2	.44				.67				.47		-.40		.69						.24
Flu3	.59					1.05			.65			.59	.54		.89				
Fle3	.52					.83			.56			.57	.45		.71				
Ori3	.31					.50			.33			.37	.33		.40				
Flu4	.63							1.02	.63		.35		.59					.82	
Fle4	.52							.73	.53		.34		.42					.63	
Ori4	.47							.60	.49		.26		.46					.46	
Flu5	.70		.90						.63		.46		.61			.72			
Fle5	.61		.90						.55		.43		.53			.72			
Ori5	.53		.76						.49		.44		.47			.58			
Flu6	.62			.94					.49	.80			.54	.79					
Fle6	.60			.99					.47	.83			.53	.83					
Ori6	.43			.52					.34	.45			.41	.40					
Eigenvalue	6.05		6.05	2.17	1.94	1.59	1.35	1.15	6.05	2.17	1.94	1.59	6.05	2.17	1.94	1.59	1.35	1.15	0.67
% Variance	29.88		12.48	12.15	12.07	11.41	11.24	10.92	30.18	9.66	8.95	8.10	26.84	8.71	8.46	8.31	7.94	7.66	6.63
Factor correlations		F1	—																
		F2	.44	—															
		F3	.24	.27	—														
		F4	.28	.23	.40	—													
		F5	.34	.33	.40	.21	—												
		F6	.41	.30	.26	.38	.30	—											

Table 1 (*continued*)

Indicators		One-factor solution	Six-factor solution						Three-bifactor solution				Six-bifactor solution						
			F1	F2	F3	F4	F5	F6	G	F1	F2	F3	G	F1	F2	F3	F4	F5	F6
Fit indices	χ^2	2779.47	83.87						1305.79				52.20						
	df	135	60						87				48						
	p	< .0001	.02						< .0001				.31						
	χ^2/df	20.59	1.40						15.01				1.09						
	CFI	.33	.99						.69				.999						
	TLI	.24	.98						.45				.997						
	RMSEA	.26	.04						.22				.02						
	SRMR	.15	.01						.07				.01						
	BIC	2005.03	-26`0.32						806.71				-223.16						
	SABIC	2433.20	-70.02						1082.64				-70.92						

Note. *n* = 310; Flu = fluency; Fle = flexibility; Ori = originality; F = factor; G = general factor; CFI = comparative fit index; TLI = Tucker Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; BIC = Bayesian information criterion; SABIC = sample-size adjusted BIC. The values of likelihood ratio test and Akaike information criterion were not available. *Factor loadings* below *.20* were not reported.

Table 2

Results of the Classical CFA

Indicators	One-factor model		Three-factor model				Six-factor model						Six-bifactor model						
			Flu	Fle	Ori		Act1	Act2	Act3	Act4	Act5	Act6	G	Act1	Act2	Act3	Act4	Act5	Act6
Flu1	.60		.58				.98						.56	.83					
Fle1	.59			.50			.88						.54	.67					
Ori1	.48				.44		.61						.42	.44					
Flu2	.71		.67					1.02					.61		.84				
Fle2	.66			.57				.87					.56		.65				
Ori2	.57				.56			.75					.46		.57				
Flu3	.64		.64						1.02				.58			.92			
Fle3	.60			.58					.88				.54			.64			
Ori3	.42				.41				.54				.35			.37			
Flu4	.68		.60							1.06			.66				.88		
Fle4	.48			.43						.75			.42				.58		
Ori4	.58				.59					.65			.55				.38		
Flu5	.62		.52								1.01		.61					.78	
Fle5	.48			.37							.80		.42					.71	
Ori5	.57				.55						.70		.54					.49	
Flu6	.53		.52									1.00	.54						.83
Fle6	.49			.52								.94	.50						.80
Ori6	.42				.42							.54	.41						.37
Factor correlations		Flu	—			Act1	—												
		Fle	1.32	—		Act2	.34	—											
		Ori	1.12	1.08	—	Act3	.29	.48	—										
						Act4	.34	.34	.30	—									
						Act5	.39	.26	.30	.39	—								
						Act6	.36	.23	.27	.38	.41	—							

Table 2 (*continued*)

Indicators		One-factor model	Three-factor model			Six-factor model						Six-bifactor model						
			Flu	Fle	Ori	Act1	Act2	Act3	Act4	Act5	Act6	G	Act1	Act2	Act3	Act4	Act5	Act6
Fit indices	χ^2	2752.89	2482.56			205.33						197.37						
	df	135	132			120						117						
	p	< .0001	< .0001			< .0001						< .0001						
	χ^2/df	20.39	18.81			1.71						1.69						
	CFI	.35	.42			.98						.98						
	TLI	.27	.33			.97						.97						
	RMSEA	.26	.24			.05						.05						
	SRMR	.14	.14			.06						.05						
	LR	-11410.64	-11275.48			-10136.86						-10132.88						
	AIC	22893.28	22628.95			20375.72						20373.76						
	BIC	23026.37	22773.14			20564.27						20573.40						
	SABIC	22912.20	22649.45			20402.53						20402.14						

Note. *n* = 311; Flu = fluency; Fle = flexibility; Ori = originality; Act = activity; G = general factor; CFI = comparative fit index; TLI = Tucker Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; LR = likelihood ratio test; AIC = Akaike information criterion; BIC = Bayesian information criterion; SABIC = sample-size adjusted BIC.

Table 3

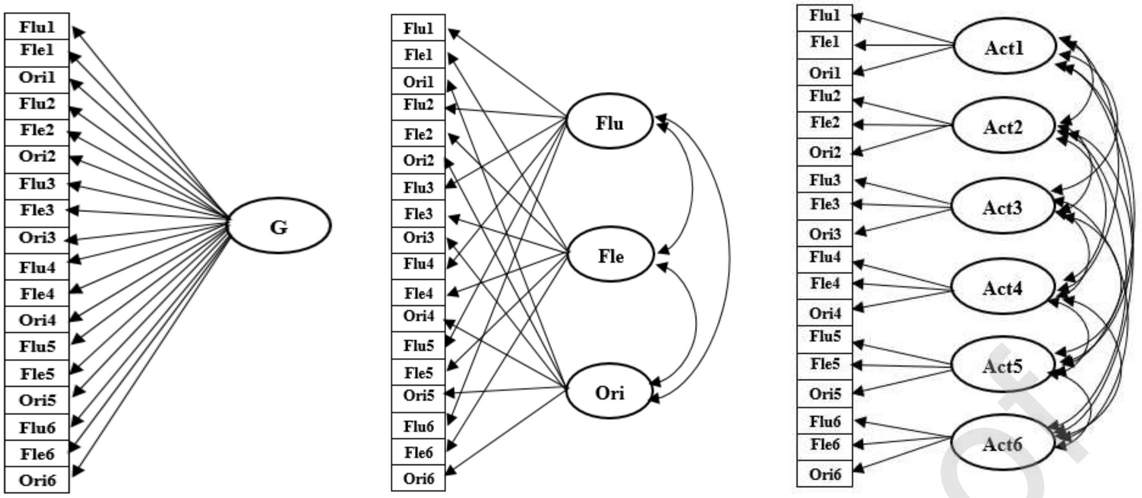
Results of the MIRT Analysis

Indicators	One-factor model		Three-factor model				Six-factor model						Three-bifactor model				Six-bifactor model						
			Flu	Fle	Ori		Act1	Act2	Act3	Act4	Act5	Act6	G	Flu	Fle	Ori	G	Act1	Act2	Act3	Act4	Act5	Act6
Flu1	.66		.65				.97						.58	.27			.57	.80					
Fle1	.64			.64			.92						.56		.20		.56	.72					
Ori1	.54				.53		.63						.45			.14	.46	.45					
Flu2	.74		.71					.99					.91	.40			.66		.75				
Fle2	.69			.66				.93					.87		.48		.60		.72				
Ori2	.57				.55			.73					.68			.15	.49		.56				
Flu3	.69		.68						.99				.64	.06			.62			.78			
Fle3	.65			.65					.95				.61		.11		.58			.75			
Ori3	.40				.39				.50				.36			.05	.35			.37			
Flu4	.65		.64							.99			.58	.36			.62				.78		
Fle4	.49			.48						.87			.42		.29		.44				.76		
Ori4	.54				.54					.66			.49			.09	.55				.42		
Flu5	.61		.61								.97		.51	.46			.59					.73	
Fle5	.49			.48							.88		.42		.26		.42					.85	
Ori5	.56				.57						.71		.47			.21	.53					.52	
Flu6	.55		.55									.99	.47	.41			.51						.86
Fle6	.50			.51								.97	.43		.33		.48						.84
Ori6	.43				.44							.56	.35			.63	.42						.40
Factor correlations		Flu	—			Act1	—																
		Fle	.99	—		Act2	.39	—															
		Ori	.97	.97	—	Act3	.32	..53	—														
						Act4	.37	.35	.32	—													
						Act5	.35	.27	.31	.39	—												
						Act6	.31	.20	.26	.36	.37	—											

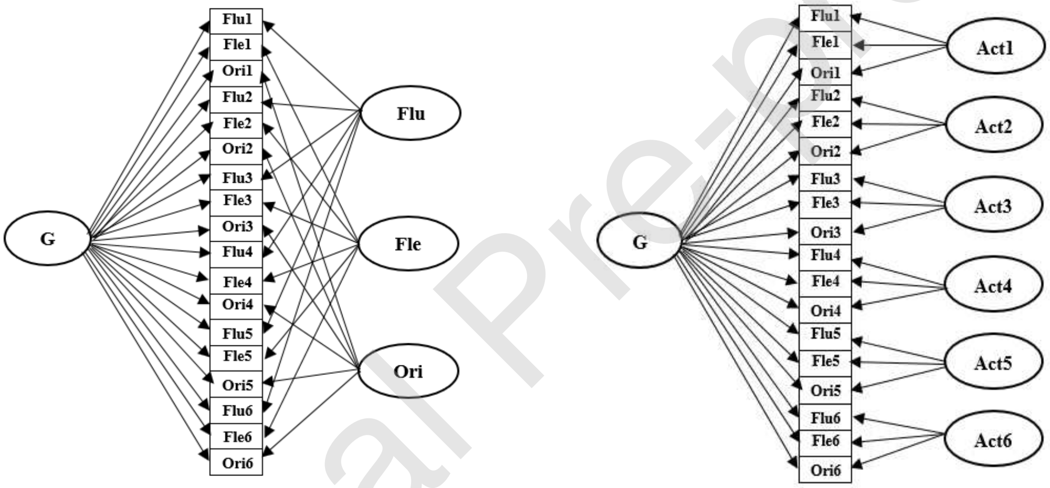
Table 3 (*continued*)

Indicators		One-factor model	Three-factor model			Six-factor model						Three-bifactor model				Six-bifactor model						
			Flu	Fle	Ori	Act1	Act2	Act3	Act4	Act5	Act6	G	Flu	Fle	Ori	G	Act1	Act2	Act3	Act4	Act5	Act6
Fit indices	C2	2007.60	2542.83			881.13						1622.66					165.05					
	df	135	132			120						117					117					
	p	< .0001	< .0001			< .0001						< .0001					.002					
	CFI	.66	.56			.86						.72					.99					
	TLI	.61	.49			.82						.64					.99					
	RMSEA	.22	.25			.15						.21					.04					
	SRMR	.15	.32			.26						.16					.05					
	LLR	-10894.97	-10901.49			-10177.08						-10792.30					-9587.35					
	AIC	22271.93	22290.97			20866.16						22102.60					19692.70					
	BIC	23173.22	23203.48			21823.55						23071.21					20661.30					
	SABIC	22408.86	22429.60			21011.61						22249.75					19839.85					

Note. n = 311; Flu = fluency; Fle = flexibility; Ori = originality; Act = activity; G = general factor; CFI = comparative fit index; TLI = Tucker Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; LLR = log-likelihood ratio test; AIC = Akaike information criterion; BIC = Bayesian information criterion; SABIC = sample-size adjusted BIC.



One-factor model **Three-factor model** **Six-factor model**



Three-bifactor model **Six-bifactor model**

Figure 1. Models of the Verbal TTCT tested in this study
Note. Flu = fluency; Fle = flexibility; Ori = originality; G = general factor; Act = activity.

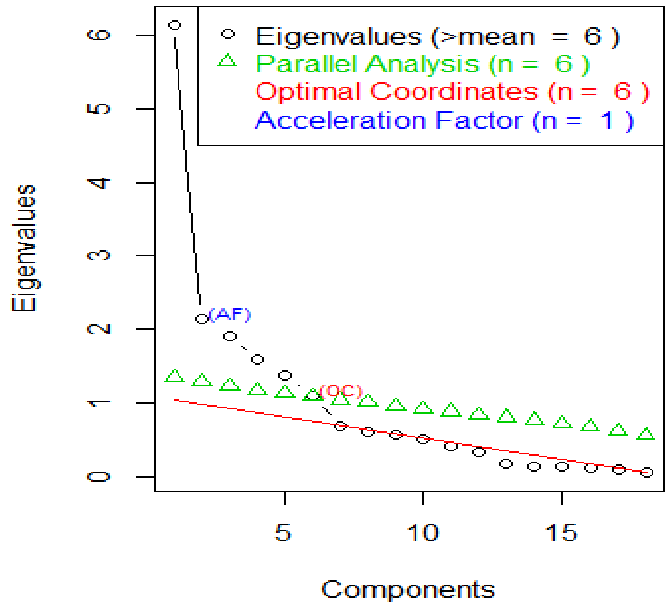


Figure 2. Non-graphical solutions to the scree test