The European Proceedings of
Social & Behavioural Sciences
EpSBS

Future Academy ISSN: 2357-1330

https://dx.doi.org/10.15405/epsbs.2018.11.02.80

ICPE 2018
International Conference on Psychology and Education

ASSESSING CREATIVE EXPRESSIVENESS IN CHILDREN'S WRITTEN STORIES USING THE CONSENSUAL ASSESSMENT TECHNIQUE

Toivainen T. (a), I. Badini (a), R. Chapman (a), M. Malanchini (b), B. R. Oliver (a),
D. Matsepuro (c), Y. Kovas (a, c)*
*Corresponding author

(a) Department of Psychology, Goldsmiths, University of London, UK
(b) Department of Psychology, Population Research Center, University of Texas at Austin, U.S.
(c) Tomsk State University, Tomsk, Russia, y.kovas@gold.ac.uk

Abstract

The study investigated methodological issues relating to the use of the Consensual Assessment Technique (CAT) for measuring creativity in children's written stories. The CAT is a commonly used measure to estimate creativity of a product, based on social recognition of creativity by independent judges. Across domains, the CAT has shown high inter-rater reliability. The present study utilised the CAT to assess creativity in children's written stories. The stories were also evaluated for: Imagination, Novelty, Liking (how much the judges liked the story), Detail, Emotion, Vocabulary, Straightforwardness, Logic and Grammar. The sample consisted of 277 nine-year-olds. The results showed that to reach sufficient inter-rater reliability, 5 coders were needed. The results gave evidence of a 2-factor structure among the 10 dimensions, indexing 'Creative Expressiveness' and 'Logic' constructs related to individual differences in writing. Girls outperformed boys on both constructs. The story length was positively correlated with the constructs, explaining 63% of the variance in Creative Expressiveness, and 42% in Logic. Creative Expressiveness was positively correlated with verbal ability (r = .20) and with teacher rating of writing (r = .28). Similarly, Logic was also correlated with verbal ability (r = .34) and teacher rating of writing (r = .44). The findings inform future research employing the CAT to measure creativity in children's storytelling.

Keywords: Creativity, Consensual assessment technique, Children's writing.

https://dx.doi.org/10.15405/epsbs.2018.11.02.80
Corresponding Author: Y. Kovas
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

1. Introduction

The Consensual Assessment Technique (CAT) is used to operationalize the creativity of a product (Amabile, 1982; Hennessey & Amabile, 2010). In the last decades, the CAT has been widely used in creativity research. For example, the CAT has been used to assess creativity in different artistic and verbal outputs as well as performance in problem solving tasks (Hennessey, Amabile, & Muller, 2011). The use of the CAT has demonstrated that people can recognise and agree upon creativity even though it may be difficult to define and characterise (Hennessey & Amabile, 2010). The CAT is based on the idea that creativity is dependent on social recognition; a product or response is considered creative to the extent that independent observers agree that it is creative (Amabile, 1982) .The CAT involves a group of independent judges, with some familiarity with the domain to which the product belongs, subjectively evaluating the creativity of a product (Hennessey, Amabile, & Muller, 2011). Also, the assessed products should be presented in a random order to the coders and they should be assessed in relation to each other, in a restricted sample of products (Hennessey, Amabile, & Muller, 2011). Due to its simplicity and consistency, the CAT has been regarded as particularly suitable to evaluate everyday creative outputs (Runco, 2004). With wide applicability, the CAT is commonly used in creativity research (Hennessey & Amabile, 2010).

In children, the CAT has been used to evaluate creativity of musical compositions, drawings and poems (Hickey, 2001; Baer, Kaufman, & Gentile, 2004; Lubart, Pacteau, Jacquet, & Caroff, 2010). Three previous studies have utilised the CAT to estimate creativity in children’s orally told or written stories (Hennessey & Amabile, 1988; Toivainen et al., 2017; Badini et al., in press). The first study established the use of the CAT in children’s stories and investigated the relationship of objective story features to creativity (Hennessey & Amabile, 1988). The study reported positive correlations between creativity and the story length (r = .28); inclusion of dialogue (r = .46); and whether the children had named the characters (r = .35). Age (range 5 – 10 years) and sex were not associated with creativity (Hennessey & Amabile, 1988). However, the study did not report the distributions of either sex or age, so further investigations into their potential role in childhood creativity are needed. A recent pilot study, investigating the relationship between creativity in writing and further educational achievement, ran an exploratory principal component analysis among 10 dimensions (see below; Toivainen et al., 2017). A summed component score, termed ‘Creative Expressiveness’, was based on 7 of the 10 dimensions that had high loadings on the principal component (Toivainen et al., 2017). This study found that the Creative Expressiveness score explained an additional 7% of variance in English exam performance at age 16, beyond intelligence and English grade at age 9 (Toivainen et al., 2017). Another recent study (based on the same sample as the present study) investigated early cognitive predictors of creativity in writing and reported a weak but significant association between early drawing ability and Creativity Expressiveness in writing at age 9 (r = .17; Badini et al., in press).

In the aforementioned three studies, the stories were coded for 10 dimensions: 1) Creativity; 2) Imagination; 3) Novelty; 4) Liking; 5) Detail; 6) Emotion; 7) Vocabulary; 8) Straightforwardness; 9) Logic; and 10) Grammar. The first study utilising these dimensions to assess children’s orally told stories, found support for a 3-factorial structure (Hennessey & Amabile, 1988). The first factor had high loadings of Creativity, Liking, Novelty and Imagination; the second of Detail and Straightforwardness; and the third of Grammar and Logic dimensions (Hennessey & Amabile, 1988). Vocabulary and Emotion dimensions did not load on any of the three factors (Hennessey & Amabile, 1988). However, only 30 out of 115 stories were coded for all 10 dimensions, as the focus of this study was on the Creativity dimension (Hennessey Amabile, 1988).

Two recent studies that assessed the 10 dimensions gave support for a 2-factorial structure (Toivainen et al., 2017; Badini et al., in press). The first factor (Creative Expressiveness) had high loadings from the following seven dimensions: Creativity; Imagination; Novelty; Liking; Detail; Emotion; and Vocabulary. The remaining three dimensions of Straightforwardness; Logic; and Grammar loaded on the second factor (Logic). In summary, previous studies have shown that Creativity loads on the same factor with Imagination, Novelty and Liking (Hennessey & Amabile, 1988), as well as with Detail, Emotion and Vocabulary (Toivainen et al., 2017). Based on this multidimensionality, the composite score was named as Creative Expressiveness to capture all dimensions that were associated with creativity in children's storytelling (Toivainen et al., 2017).

More research is needed into associations between creativity and domain specific- and domain-general abilities, which are prerequisites for creative outputs (Amabile, 1983). A pilot study on creativity in writing and later educational achievement found no significant correlation between Creative Expressiveness scores and general cognitive ability at age 9 (Toivainen et al., 2017). However, since the measure for general cognitive ability in the study was a composite of two non-verbal and two verbal measures, the specific role of verbal ability in creativity in writing was not evaluated. The same study also reported a positive correlation between Creative Expressiveness and English grade at age 9 (r = .36; Toivainen et al., 2017). Again, the English grade was a composite of teacher-reported scores of Reading; Speaking and Listening; and Writing. Further research is needed in order to evaluate the extent to which creativity in children's writing is related specifically to writing skills.

2. Problem Statement

The application of the CAT to children's creative writing needs to be further validated. In addition, research is needed into inter-relationship between different dimensions of written stories assessed by the CAT, and into associations between creativity and specific abilities, such as verbal ability and writing skills. Also, research on children's writing has not explored so far the relationship between the story length and creativity, which is relevant due to the variability in the lengths in writing tasks with no word limits (21 to 486 words in this sample). Furthermore, the question of sex differences in creativity in childhood writing is still unanswered.

3. Research Questions

1. How many coders are needed to reach sufficient inter-rater reliabilities on the 10 dimensions of the CAT?

2. Are the 10 dimensions correlated, and to what extent?

3. Does confirmatory factor analysis support2-factorial solution among the 10 dimensions, indicated in the previous pilot study?

4. Are there gender differences in factor scores?

5. Does the story length correlate with the factor scores? Is the association similar at different levels of the story lengths?

6. Are the factor scores correlated with verbal ability and teacher rating for writing at age 9?

4. Purpose of the Study

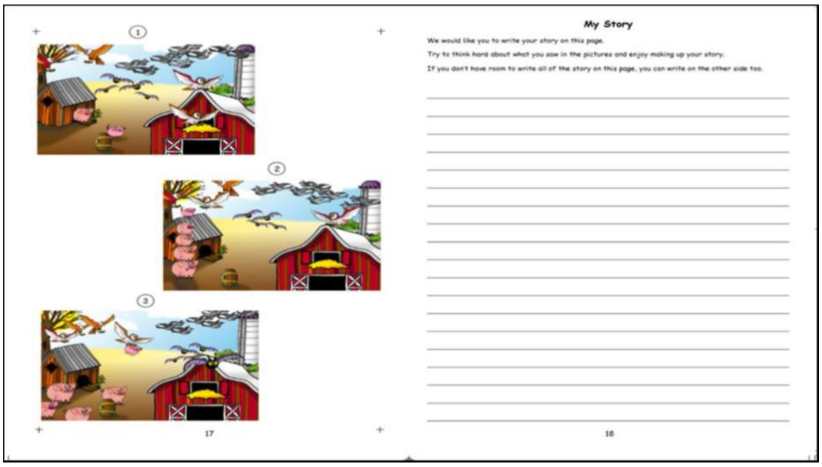
The present study seeks to fill a gap in the literature by investigating in detail the suitability and potential methodological issues of using all 10 dimensions of the CAT in the assessment of creativity in children’s written stories. The results of this study will inform a planned future large-scale, genetically informed study (n = 1300) using the same measure. It is important to establish the validity of the CAT before coding more stories as the coding procedure is very intensive. The procedure requires transcribing and reading all the stories in a sample before coding commences. The findings will provide new insights into creativity in writing and will further evaluate construct ‘Creative Expressiveness’ that was suggested by a previous study (Toivainen et al., 2017).

5. Research Methods

The sample used in the present study is a subsample from the Twins Early Development Study (TEDS). TEDS is a large, longitudinal twin sample that includes more than 13,000 twin pairs, born between 1994 and 1996, representative of the population of England and Wales (Haworth et al., 2013). The total sample in the present study was 277 with a mean age of 9.02 years (SD = .27), ranging from 8.50 to 9.82 years. Age was recorded at the time when test booklets were returned. Only one twin per pair was selected, in order to eliminate the inflated inter-individual similarity observed in twins. The sample consisted 172 girls (Mage = 9.02, SD = .28) and 105 boys (Mage 9.02, SD = .27). The present study is part of a larger longitudinal study, which focuses on measures at ages 4, 9 and 16, so preliminary sample selection was conducted among participants who had provided data at all three collection waves. Only data from the collection at age 9 was used in the current study. Preliminary analyses were run to establish the representativeness of the selected subsample. In the present study, the mean for verbal ability was slightly higher (M = .06, SD = .98) than for the whole TEDS sample, which is representative of the population of England and Wales and has a standardised mean of 0. Further, the teacher rated scores of writing were slightly higher (M = 3.01, SD = .68) in comparison with the larger TEDS sample (M = 2.83, SD = .74).

5.1. Written stories at age 9

The children were shown three coloured pictures of farm animals and farm buildings. They were then instructed to write a story that was creative. The pictures and instructions for the task are shown below in Figure 01. The data were collected in children’s homes. The stories were written in 2002-2004. There was no time limit for the task and it was instructed and supervised by the parents/guardians of the children. The stories were first transcribed to minimise the influence of differences in handwriting on coding. No corrections were made to spelling, grammar etc. during transcription. The length of the stories ranged from 21 to 486 words, with a mean of 147.99 (SD = 80.55) words.



My Story

We would like you to write your story on this page.

Try to think hard about what you saw in the pictures and enjoy making up your story.

If you don't have room to write all of the story on this page, you can write on the other side too.

Figure 01. The pictures and instructions for the 'My Story' task

The stories were coded for the following 10 dimensions: 1) Creativity; 2) Imagination; 3) Novelty; 4) Liking; 5) Detail; 6) Emotion;7) Vocabulary; 8) Straightforwardness; 9) Logic; and 10) Grammar. Five independent judges coded the stories for these 10 dimensions, each on a 7-point Likert-scale using their own subjective interpretation of each dimension. For example, for the creativity dimension, the judges were instructed as follows: "Please evaluate the creativity of the story on this page in relation to the other 276 stories. Use your own subjective assessment of creativity". No other criteria and instructions were given. Firstly, all the judges were asked to code the stories only for creativity. After coding all the stories for creativity, the judges were asked to then code them for the remaining nine dimensions. For these dimensions the judges were asked to again use their subjective assessments (e.g. "Please evaluate the straightforwardness of the story on this page in relation to the other 276 stories. Use your own subjective assessment of straightforwardness."). The stories, and additional 9 coding dimensions, were presented to the judges in different orders to counterbalance for potential order effects. The judges were adults, primarily undergraduate psychology students.

5.2. Verbal ability and teacher ratings for writing, as measured at age 9

Verbal ability at age 9 was assessed using vocabulary and general knowledge tests adapted from the WISC-III-UK (Wechsler, 1992; e.g. Vocabulary: 'What does migrate mean?'; General Knowledge: 'In which direction does the sun set?'). The total score was a composite of the two tasks scores. The score for English writing was a single teacher-reported subscore of English score (the other subscores were reading; and speaking & listening). Teachers were asked to evaluate children's writing attainment (scale 1-5) in terms of the National Curriculum. Score 1 represented writing attainment well below the expected standard for most 9-year-olds, whereas score 5 was an indicator of exceptional achievement in writing, above the level expected at age 9.

6. Findings

6.1. How many coders are needed to reach sufficient reliabilities in the 10 dimensions?

Table 1 presents the increments of internal reliabilities for each dimension from 2 to 5 coders. For 7 dimensions (Creativity; Imagination; Novelty; Liking; Detail; Emotion; and Vocabulary), the reliabilities exceeded the recommended minimum α = 0.70 with 2 coders (Nunnally & Bernstein, 1994). With 5 coders, 9 dimensions had internal reliabilities higher than α = 0.70. Cronbach's alpha for Straightforwardness was 0.67.

Table 01. Internal reliability (Cronbach's α) for the 10 coding dimensions as a function of the number of the coders

Dimension	2 coders	3 coders	4 coders	5 coders	Δ
1. CR	.79	.85	.86	.88	.09
2. IM	.78	.81	.84	.86	.08
3. NO	.79	.82	.83	.85	.06
4. LI	.76	.79	.82	.84	.08
5. DE	.78	.79	.83	.86	.08
6. EM	.78	.82	.83	.86	.08
7. VO	.74	.78	.81	.85	.11
8. ST	.20	.40	.56	.67	.47
9. LO	.43	.56	.66	.73	.30
10. GR	.66	.69	.72	.77	.11

Note. n = 277; CR = Creativity; IM = Imagination; NO = Novelty; LI = Liking; DE = Detail; EM = Emotion; VO = Vocabulary; ST = Straightforwardness; LO = Logic; GR = Grammar; Δ = increment in α, between 2 and 5 coders.

6.2. What are the correlations between the 10 dimensions?

The bivariate correlation coefficients between the 10 dimensions are shown in the Table 2.

Table 02. Bivariate correlations between the 10 coding dimensions

	1. CR	2. IM	3. NO	4. LI	5. DE	6. EM	7. VO	8. ST	9. LO	10. GR
1. CR	1									
2. IM	.89	1								
3. NO	.85	.87	1							
4. LI	.83	.83	.82	1						
5. DE	.74	.73	.68	.73	1					
6. EM	.73	.73	.69	.74	.66	1				
7. VO	.66	.64	.58	.68	.70	.66	1			
8. ST	.23	.23	.25	.35	.27	.27	.35	1		
9. LO	.28	.25	.26	.42	.32	.28	.37	.68	1	
10. GR	.14	.14	.12	.19	.21	.22	.28	.34	.26	1

Note. n = 1385; All correlations are significant p < .001; CR = Creativity; IM = Imagination; NO = Novelty; LI = Liking; DE = Detail; EM = Emotion; VO = Vocabulary; ST = Straightforwardness; LO = Logic; GR = Grammar.

Most of the zero-order, bivariate correlations between the 10 dimensions were moderate to high. The inter-correlations between Creativity, Imagination, Novelty and Liking were higher than r = .82. The last three dimensions (Logic, Straightforwardness and Grammar) had lower bivariate correlations with the other 7 dimensions (highest correlation r = .42). Logic and Straightforwardness were correlated at r = .68.

6.3. Does confirmatory factor analysis support the 2-factorial solution among the 10 dimensions?

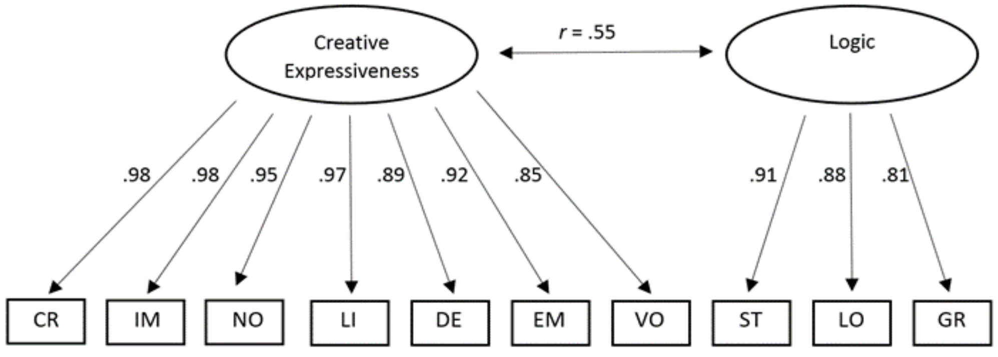
Previous pilot study using the CAT with 10 dimensions for assessment of creativity in children's written stories has suggested a 2-factorial structure (Toivainen et al., 2017). Confirmatory factor analyses (CFA) were run to test if the 2-factorial model fits the data better than a model in which all dimensions load onto a single factor. 3-factorial model, as indicated by Hennessey & Amabile (1988) was inadmissible due to the high correlations between the three latent factors and therefore the fit indices for 2-factorial model were compared with a 1-factorial model. The model fit outputs for 1 and 2-factorial models are presented in Table 3.

Table 03. Confirmatory factor analyses fit indices for 1-factor and 2-factor solutions for the 10 coding dimensions

Model	AIC	BIC	X^2	RMSEA	CFI	TLI	SRMR
2-factorial	13263.28	13339.31	488.93*	0.22	.90	.86	0.09
1-factorial	13640.01	13712.43	867.67*	0.29	.81	.76	0.13

Note. * p < .001; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index = SRMR = Standardised Root Mean Square Residual

As shown, a 2-factor model is a better fit for the data than a 1-factor model. This is indicated by the lower X2, as well as lower AIC and BIC indices; higher CFI and TLI values; and lower values of RMSEA and SRMR. The factor loadings for the 2-factor model are presented in Figure 2.



Note. CR = Creativity; IM = Imagination; NO = Novelty; LI = Liking; DE = Detail; EM = Emotion; VO = Vocabulary; ST = Straightforwardness; LO = Logic; GR = Grammar.

Figure 02. Factor loadings (and the correlation between the latent variables) for 2-factor solution for the 10 coding dimensions

Based on the results of the CFA, the scores for these two factors were created by combining the scores from each five judges for each dimension that had high loadings on each factor. The summed scores were used, as opposed to weighted values, due to the small differences in factor loadings on each factor (in Creative Expressiveness .85 - .98; in Logic .81 - .91). The mean total factor scores, based on the scores from 5 coders, for Creative Expressiveness (factor score) is 105.51 (SD=34.48) and for Logic (factor score) 65.67 (SD= 12.19). The two factors have different numbers of dimensions and therefore widely different means. The difference in means do not affect any of the analyses.

6.4. Are there gender differences in Creative Expressiveness and Logic factor scores?

For Creative Expressiveness, the mean difference between girls (M= 110.80, SD= 33.81) and boys (M= 96.84, SD= 33.96) was significant (t (276) = 3.33, p< .01; d = .41). Girls (M = 67.09, SD = 11.99) also outperformed boys (M = 63.34, SD = 12.22) in Logic vs.; t (276) = 2.51, p = .01; d = .31).

6.5. Does number of words correlate with Creative Expressiveness and/or Logic Factor scores?

The mean story length was 148 words (SD = 80.55). The lengths varied between 21 and 486 words. The number of words in a story had positive correlations with both Creative Expressiveness and Logic. Linear regression analyses showed that the number of words accounted for 63.2% of the variance in Creative Expressiveness and 17.4% in Logic.

Quantile regressions were run to establish if the associations between story length and factor scores (Creative Expressiveness and Logic) were similar at different levels of story length. The stories consisted of 21-91 words in the first quantile (n=70); 93-132 words in the second quantile (n=69); 133-178 words in the third quantile (n=69); and 181-486 words in the fourth quantile (n=69). The beta coefficients were similar for both measures in all 4 quantiles. Intercepts increased in-line with quantiles, indicating that the associations between the story length and factor scores, for both Creative Expressiveness and Logic, are similar in all 4 length quantiles.

Table 04. Intercepts and beta coefficients for 4 quantiles of Story Length (number of words) predicting Creative Expressiveness

Quantiles for Story Length	Intercept	Beta co-efficient	Confidence interval	t-value
1st	35.71	0.36	[.30, .40]	11.19*
2nd	40.69	0.40	[.36, .43]	18.53*
3rd	53.03	0.38	[.35, .43]	21.99*
4th	63.98	0.39	[.33, .50]	15.09*
Total	55.15	0.34	[31., .37]	21.74*

* p < .01

Table 05. Intercepts and beta coefficients for 4 quantiles of Story Length (number of words) predicting Logic

Quantiles for Story Length	Intercept	Beat co-efficient	Confidence Interval	t-value
1st	47.25	.07	[.04, .09]	5.89*
2nd	54.40	.07	[.04, .08]	5.25*
3rd	60.45	.06	[.04, .09]	7.18*
4th	66.29	.06	[.04, .08]	5.14*
Total	56.06	.06	[.05, .08]	7.62*

* p < .01

6.6. Are the factor scores correlated with verbal ability and teacher rating for writing at age 9?

Creative Expressiveness and Logic were both positively correlated with verbal ability and teacher rating for writing, as measured at age 9. As seen in Table 5, the correlations for both verbal ability and teacher rated writing were stronger for Logic than for Creative Expressiveness.

Table 06. Bivariate correlations for Creative Expressiveness; Logic; verbal ability at 9; and teacher rating for writing at 9

	1. Creative Expressiveness	2. Logic	3. Verbal ability at 9	4. Teacher rating for writing at 9
1.	1			
2.	.55	1		
3.	.20	.34	1	
4.	.28	.44	.37	1

Note. n = 277; All correlations are significant at p < .01

7. Conclusion

The present study investigated the use of the Consensual Assessment Technique (CAT) for assessing creativity in children’s written stories. Creativity dimension was studied in relation to nine other dimensions: Imagination, Novelty, Liking, Detail, Emotion, Vocabulary, Straightforwardness, Logic and Grammar. Firstly, we established the number of judges needed to reach sufficient inter-rater reliabilities for the 10 coding dimensions. Secondly, we examined the correlations between the 10 dimensions and replicated the previously established 2-factor structure among the 10 dimensions. Thirdly, we explored how Creative Expressiveness and Logic factor scores relate to gender; story length; verbal ability; and teacher rated English writing score.

Our results showed that five coders are needed to reach sufficient inter-rater reliability levels for all dimensions except for Straightforwardness, for which the level of inter-rater reliability was lower (.67) than the recommended α = .70 (Nunnally & Bernstein, 1994). The lower inter-rater reliability in Straightforwardness may reflect different interpretations of the dimension among the judges. The scoring was based on the coders’ subjective evaluations and not on any objective criteria. Rating 277 stories required a substantial time commitment from each coder. Moreover, reliability increments for several dimensions were small when number of coders increased. This suggests that 5 coders would be optimal for future uses of the CAT to evaluate 10 dimensions of children's writing. Factor scores were calculated as summed scores from each coder, based on the highest loading dimensions. The dimensions included in Creative Expressiveness were: Creativity, Imagination, Novelty, Liking, Detail, Emotion, and Vocabulary. The Logic factor score was comprised of the sum of scores from the Straightforwardness; Logic; and Grammar dimensions.

All the story dimensions were inter-correlated. Confirmatory factor analysis supported a 2-factor structure suggested by an exploratory factor analysis of the previous pilot study (Toivainen et al. 2017). The seminal study, which established the use of CAT for evaluation of creativity in children's storytelling, reported a 3-factorial model based on the 10 coding dimensions (Hennessey & Amabile, 1988). The difference with the factor structure found in the present study may be due to differences in data collection (oral vs. written stories). It is plausible that when children are telling stories aloud, it is easier for them to be more detailed and elaborate. Hand-written stories require additional skills not needed for oral stories, such as fine-tuned motor skills. Also, interest and enjoyment in writing is likely to influence the amount of time children are spending on the task. Participants in the earlier study also had a wider age range, 5 to 10 years, whereas the children taking part in the present study were 9-years-old. These reasons may have influenced the content of the stories and subsequently how they were scored on the 10 dimensions. Additionally, the present study used a bigger sample than the previous study in which only 30 stories were coded for all 10 dimensions.

The finding that the Logic score had a stronger positive correlation with verbal ability and teacher rating for writing reflects the dimensions that constitute the Logic Factor score: Straightforwardness, Logic and Grammar; each of which is related to logical reasoning. The scoring on these items may have emphasised technical writing skills. Verbal ability is measured by verbal reasoning tasks and teachers emphasise technical writing skills over creative expression when assessing nine-year-olds' writing skills. Therefore, several dimensions that are included in Creative Expressiveness, such as Imagination and Emotion would not be reflected in either verbal ability or in teacher rated writing scores.

Further studies on creativity in children's stories should take into consideration the role of gender and length of the stories. At age 9, girls scored higher than boys in both Creative Expressiveness (d = .41) and Logic (d = .31) factors. This result is in-line with previous research that has shown that girls outperform boys in writing at age 9 (Kovas, Haworth, Dale & Plomin, 2007). The results also showed a substantial, positive correlation between the story length and Creative Expressiveness. It is likely that shorter stories do not allow for much creative expression, for example through a sophisticated narrative structure. This may be particularly relevant in children's writing as nine-year-olds have a limited vocabulary and experience of different forms of writing. The associations between number of words and creativity were similar at different levels of Creative Expressiveness; among the shortest stories (the first quantile; i.e. fewer than 91 words) story length was still associated positively with creativity. Similarly, among the longest stories (the fourth quantile; more than 181 words), shorter ones were evaluated as being less creative.

The results of the study contribute to research on valid and reliable methods of assessing individual differences in creativity among children. These methods will improve the quality of research into aetiology of individual differences in creativity; and can be used as an educational diagnostic tool.

Acknowledgments

We gratefully acknowledge the on-going contribution of the participants in the Twins Early Development Study (TEDS) and their families. TT is supported by the Economic and Social Research Council [grant number ES/J500124/1]. This work was supported by the Russian Ministry for Education and Science (project No. 25.8562.2017/9.10).

References

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*(5), 997-1013.

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology*, *45*(2), 357.

Badini, I., Toivainen, Malanchini, M., Oliver, B. R., & Kovas, Y. (in press).Early Human Figure Drawing as a Predictor of Creative Expressiveness in Childhood. In *ICPE: The European Proceedings of Social and Behavioural Science*.

Baer, J., Kaufman, J. C., & Gentile, C. A. (2004). Extension of the consensual assessment technique to nonparallel creative products. *Creativity research journal*, *16*(1), 113-117.

Hennessey, B. A., & Amabile, T. M. (1988). Story-telling: A method for assessing children's creativity. *The Journal of Creative Behavior*, *22*(4), 235-246.

Hennessey B. A., & Amabile, T.M. Creativity (2010). *Annual Review of Psychology*, *61*, 569-98.

Hennessey, B. A., Amabile, T. M., and Mueller, J. S. (2011). Consensual Assessment. In M. A. Runco, S. R. Pritzker (Eds.), *Encyclopedia of Creativity* (2nd ed.) pp. 253-260. San Diego, US: Academic Press.

Hickey, M. (2001). An application of Amabile's consensual assessment technique for rating the creativity of children's musical compositions. *Journal of Research in Music Education*, *49*(3), 234-244.

Lubart, T., Pacteau, C., Jacquet, A. Y., & Caroff, X. (2010). Children's creative potential: *An empirical study of measurement issues. Learning and Individual Differences*, *20*(4), 388-392.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill

Runco, M. A. (2004). Creativity. *Annual Review of Psychology*, *55*, 657–687.

Toivainen, T., Malanchini, M., Oliver, B. R., & Kovas Y. (2017). Creative storytelling in childhood is related to exam performance at age 16. *The European Proceedings of Social & Behavioural Sciences*, *33*, 375-384. doi: 10.15405/epsbs.2017.12.40