Comparison of Competing Theories about Ideation and Creativity

Ivonne Chand O'Neal

Creativity Testing Services, Washington, DC USA

E-mail address: ichand@comcast.net

Sue Hyeon Paek

University of Georgia, Athens, GA USA

E-mail address: suehpaek@gmail.com

Mark A. Runco

University of Georgia, Athens, GA USA

E-mail address: runco@uga.edu

A R T I C L E   I N F O

A B S T R A C T

A measure of ideational behaviour, often used to estimate the potential for creative thinking, was administered to 796 children and their parents and teachers. Correlations among groups were explored. The data provided an opportunity to (a) compare four theories of creativity (a one-factor theory, 2 two-factor theories, and a three-factor theory) and (b) determine empirically how the measure of ideation should be scored (based on its empirical structure). Results of confirmatory factor analyses indicated that one of the two-factor theories (Process and Product) best fit the data and was useful for comparisons of the children and their parents and teachers. Practical implications of the differences between parents and teachers are explored. Any effort to fulfil creative potentials, for example, would probably be the most likely to succeed if children, parents, and teachers agreed, and just as probable are difficulties if the three groups disagreed or considered different things when judging creative potential. Limitations of the study are also discussed.

Potential is arguably the most important topic for empirical research on creativity. Unfortunately, creative potential is a difficult subject matter. This is because potential is quite different from actual performance. By definition, potential denotes skills and talents that are latent, as of yet unused, or perhaps immature. Performances can be observed, but potential must be inferred. No wonder most research focuses on creative accomplishments and actual creative performances. Those allow objective study.

Potential is difficult to measure, but there are ways to estimate it. Certain measures of creativity focus on potential and assess the capacity to recognize a worthwhile problem, the ability to generate numerous ideas (some of which may be original), intrinsic mo-

tivation, openness to experience, and other things which are both theoretically and empirically related to actual creative performance (Runco, 2008).

The most frequently used tests of creative potential focus on ideation and divergent thinking. These present open-ended tasks to the examinees and ask for ideas that are relevant to, or even solve, the task or problem at hand. The ideas generated are typically scored for *fluency*, which is simply the number of ideas given by any one examinee, *originality*, which is based on the statistical infrequency or unusualness of the ideas, and *flexibility*, which represents the number of conceptual categories tapped by the set of ideas given by any one individual. Several other tests of ideation seem to be useful for estimating creative potential but do not actually require the generation of ideas. The *Abedi Test*, for example, uses a multiple choice format, with every question asking something about ideas (Abedi, 2002). The *Runco Ideational Behavior Scale* (RIBS), in contrast, is a self-report Likert measure which allows individuals to describe how frequently they produce ideas, and in what situations, and what kind of ideas are in fact typically produced (Runco, Plucker & Lim, 2000-2001; Runco, Walczyk, Acar, Cowger, Simundson & Tripp, 2013).

The RIBS was actually conceived as a criterion measure. It was developed because tests of divergent thinking were not evaluated with appropriate criteria in research on their predictive validity. This research used criteria that required skills and resources outside of the realm of divergent thinking and were therefore indicative of much more than ideation (Wallach & Wing, 1969). As a result, the predictive validity of divergent thinking tests was unimpressive. This was interpreted as just another instance of the *criterion problem* that had plagued creativity research for decades (Shapiro, 1960).The RIBS was developed so that there would be a criterion focused entirely on ideas. It has proven to be useful in various studies of divergent thinking and creative potential (Chen, Roth & Todhunter, in press; Plucker, Runco & Lim, 2006; Runco et al., 2000-2001; Runco et al., 2013).

The present investigation was designed in part to examine how well children, parents and teachers agreed with one another in terms of estimates of creative potential provided by the RIBS. Such agreement, or lack thereof, would have obvious practical implications. Any effort to fulfil creative potentials, for example, would probably be the most likely to succeed if children, parents, and teachers agreed. Difficulties are probable if the three groups disagreed or considered different things when judging creative potential. Because the RIBS had never been used with these three samples, the first analyses reported below focused on the reliability and structure of the RIBS. A second research question followed from the fact that the RIBS asks about various aspects of ideation (e.g. the process

of ideation, or the products of ideation) and thus subscales are often used. This raises the question, are RIBS subscales more useful than a total RIBS score?

Four theories of creativity can be used when deciding about RIBS subscales. Each theory implies a different factor structure: one general scale using all RIBS items, a two-scale model representing the quantity and quality of ideas, another two-scale model but with product and process scales, and a three-scale model with fluency, originality and flexibility each represented. These four theories were each tested with confirmatory factor analysis (CFA). This analysis was useful as both a means for refining the RIBS and as an empirical test of four theories of creativity. For clarity, the CFAs and test of the RIBS structure are reported as Study 1, below. These analyses were useful in terms of refining and understanding the RIBS but also facilitated the subsequent analyses, which compared the various samples of participants.

Study 2 used the RIBS models that had the best fit to compare the parent and teacher samples. Were the scales of the RIBS useful in both groups? Was there invariance in the samples? This is one way to get at agreement or disagreement among the parents and teachers when assessing children's creativity.

Another way to assess the degree of agreement among samples relies on correlational analyses. Thus, after the CFAs were completed and conclusions about the RIBS scales reached, bivariate correlations were calculated to determine how the children's ratings of their own ideation were associated with the RIBS ratings given by the parents and the teachers. The correlations are reported in Study 3, again to keep the research questions distinct and for clarity. Previous research uncovered discrepancies between parents' and teachers' ratings of children (Runco, 1987) and it was quite possible that similar differences would be apparent in the RIBS ratings, in Study 3. This would say something about the best way to estimate potential and something about what kind of information (e.g. ideation) is used when parents and teachers are asked to judge the creativity of children.

STUDY 1

Study 1 focused on the data from the children and was intended to determine the most parsimonious manner for explaining performances on the RIBS (Data from parents and teachers were analyzed in Studies 2 and 3). Study 1 used structural equation modeling in order to determine the best model for the RIBS. There were four theoretically-justified models to consider. Each of them describes ideation as it relates to creativity. One of these theories holds that the three indices – fluency, originality and flexibility – are all necessary when studying or measuring divergent thinking. This theoretical approach is exemplified by Guilford (1950, 1968), although he actually preferred the term *divergent production* instead of divergent thinking. Guilford's interest was in developing tests that

produced scores that were orthogonal to one another and which helped him to understand the *structure of the intellect*. He eventually identified 180 supposedly distinct processes and skills, each of which represented one cell or component of the structure of intellect. Torrance (1974, 1995) drew from Guilford's work and developed probably the most comprehensive battery of divergent thinking tests, known as the *Torrance Tests of Creative Thinking,* and he almost always used these three indices. Some of his tests also used an index called *elaboration*, which represented how much detail was provided in an individual's ideational output. This theoretical approach, developed by Guilford (1968) and Torrance (1995), is labelled the divergent thinking.

A second theoretical approach to divergent thinking is just a simplified version of the three index model. This second theoretical approach focuses on two indices, namely quantity and quality (e.g. Wallach & Kogan, 1965; Wallach & Wing, 1969). You might say that the theoretical justification for this two index model is psychometric because fluency and originality do often capture the lion's share of the variability in divergent thinking test scores, and they are reliable indices, and there is research suggesting that they have moderate validity. Sometimes the two index model does not focus on fluency and originality but instead uses something like quantity and quality as the two indices. Even here, quantity is close to fluency, and quality, like originality, is often based on statistical infrequency. This approach to ideation is the *two index theory*.

A third theoretical approach to ideation and creativity also relies on two indices, but they are very different from fluency and originality, and different from quantity and quality of ideas. This third theoretical approach was suggested by Rhodes (1961) and extended by Runco (2008), Richards (1999), and Cropley and Cropley (2012). Rhodes, for example, described strands in the creativity research as falling into categories of the creative *process*, the creative *personality*, the creative *place*, or the creative *product*. These categories have proven to be enormously useful, though typically they are applied much more broadly than divergent thinking and ideation. The *place* and *person* categories do apply clearly to many aspects of creativity but would be particularly difficult to apply to ideation. The other two categories, *process* and *product*, apply to ideation very nicely, at least when ideation is assessed with the RIBS. Indeed, our own *a posteriori* examination indicated that all RIBS questions ask about an individual's ideational process, or his or her ideas that could lead to a product. This scoring method for the RIBS has recently proven to be very useful in research on the benefits of art across the disciplines (Chand O'Neal, Schulz Begle & Runco, 2014). This conceptualization of the RIBS and ideation can be summarized as the *process-product theory.*

The last theoretical approach to divergent thinking and creativity is the simplest but al-

so the weakest. It uses a *single index* and focuses entirely on fluency. As was the case with the first two index system (fluency and originality), described above, the single index approach is not really tied to a theory of creativity, nor to a theory of cognition. It too is essentially psychometric (i.e. based on sizeable correlations between fluency and originality) rather than creativity theory. It is in this sense that it is the weakest of the theories investigated here. Still, it is a very common approach to this assessment of divergent thinking (Baer, 1991; Hocevar, 1979a, 1979b, 1980; Hong, Milgram & Gorsky, 1959; Milgram & Milgram, 1976; Vartarian et al., 2014) and thus worth examining and comparing to the other theories.

Method

Participants. A total of 796 4th ($n = 407$) and 5th grade ($n = 389$) students (48.2% boys, 51.8% girls; age M = 9.68) from elementary schools in the mid-Atlantic region of the United States participated in this study. The ethnic composition of the sample was 52.1% White or Caucasian, 16.9% Asian, 13.0% Latino/Hispanic; 4.3% African American, and 13.7% mixed ethnicity. Students were sampled from two groups: One group ($n = 589$) was comprised of students enrolled in schools participating in The John F. Kennedy Center's *Changing Education Through the Arts Program* (CETA). A second group ($n = 207$) consisted of students in schools that did not participate in the CETA program.

The CETA program is a partnership between the Kennedy Center and 16 schools that have committed to a whole school focus on arts integration. Teachers participate in ongoing, sustained professional learning to build their capacity to integrate the arts throughout the curriculum. Arts integration is defined as "an approach to teaching in which students construct and demonstrate their understanding through an art form. Students engage in a creative process which connects an art form and another subject area and meets evolving objectives in both."

Both groups of participating students were matched on ethnicity, Math and Reading standardized test scores, ELL, and Title 1 status. These samples were recruited by inviting qualifying schools to participate through each school principal. Upon principal agreement, four classrooms were randomly selected in each grade level. Teachers of these classes were then invited to participate at the classroom level. Study packets which included consent and assent materials in both English and Spanish were sent home with each student. Students who returned signed consent and assent forms were included in the study. This study was approved by each of the five participating school district review boards in addition to an independent review board. Students who received parent consent but did not themselves assent were not included in the study. Data were incomplete or missing for 39 children, leaving 757 children for the analyses.

Procedure

Two experimenters and one school representative were present throughout the testing sessions at designated sites on each school campus. Students heard a script which both outlined the purpose of the study and provided instructions for completing the assessments. Provisions were made for those who required assistance due to poor sight or hearing needs. The RIBS directions were as follows:

"Use the smiley face scale (look below) to show how often each of the sentences describes you or your thinking and behaviour. Please tell us how you really think and behave, not how you would like to behave. Remember--no names are used. Your answers will be a secret. The smiley face options will help you show us how often you have done each of the 25 things on this list. For each item, circle the choice that is THE CLOSEST to being right. Here are the options:

- never

- occasionally = about once a year

- sometimes = about once or twice each month

- regularly = about once or twice each week

- very often = just about every day, and sometimes more than once each day."

The "smiley" response options had a happy face with varying degrees of happiness. The smile for "very often" was quite large, but a bit smaller for "regularly." The smile suggested ambivalence for "sometimes" and then turned to a slight and eventually a dramatic frown for "occasionally" and "never," respectively. Smiley options have proven to be useful with children in previous research (e.g. Runco & Charles, 1993).

Measure

The RIBS was developed specifically to measure ideational behaviours. Initially 100 descriptions of ideational behaviour (i.e., "behaviours" in the sense of overt action, but clearly dependent on an idea or ideas) were produced. Then redundancies were removed and factor analyses and inter-item analyses computed (Runco et al., 2000-2001). Twenty-three items were retained. Subsequent work modified the wording, and a few concepts, as needed for younger samples. This *RIBS for Children* (RIBS-C) was used in Study 1. It contains the original 23 items, reworded for children, as well as a few items which capture ideational behaviours pertaining to children's lives (e.g., I have an idea about a new way to get from home to school.)

Results

Study 1 compared four competing models in terms of best fit criteria with the RIBS data. In order to examine model fits of competing models, six model fit indices were calculated by maximum likelihood estimation with robust standard errors (MLR): chi-square,

the comparative fit index (CFI; Bentler, 1990), Tucker-Lewis Index (TLI; Bentler, 1990), Root Mean Square Error of Approximation (RMSEA; MacCallum, Browne & Sugawara, 1996), Akaike Information Criterion (AIC; Akaike, 1974) and the Bayesian Information Criterion (BIC; Schwarz, 1978).

The two-factor model A and the two-factor model B were nested within the one-factor model. However, the three-factor model was only nested within the two-factor model A and the modified two-factor model B was only nested within the two-factor model B because of item categorization. Therefore, the AIC and BIC were used to compare non-nested models and the log likelihood ratio test were used to compare nested models.

The results summarized in Table 1 suggest that the two-factor model B showed the best model fit among the four competing models. The two-factor model B showed smaller AIC and BIC than other models and the log likelihood ratio test was smaller than for the one-factor model. The model fit of the two-factor model B was adequate: RMSEA was smaller than .05 (RMSEA = .043), and CFI and TLI were close to .9 (CFI =.893 and TLI = .883), which are typical criteria of goodness-of-fit.

Table 1
Fit of Four Competing Models

	Model	CFI	TLI	RMSEA	AIC	BIC	df	χ^2	Relative χ^2 (χ^2/df)	$\Delta\chi^2$	Δdf	p
1	One-Factor Model	.873	.862	.045	57167.61	57514.81	275	701.966	2.55	-	-	-
2	Two-Factor A: Quality/Quantity	.874	.862	.045	57167.58	57519.41	274	700.249	2.56	1.72[a]	1	.19
3	Three-Factor Model	.878	.866	.045	57151.24	57512.33	272	682.889	2.63	17.36[b]	2	.001
4	Two-Factor B: Process-Product	.893	.883	.042	57089.68	57441.52	274	633.710	2.31	68.26[c]	1	NA
5	Modified Two-Factor: Process-Product	.930	.921	.034	56954.76	57348.26	265	501.569	1.89	132.14[d]	9	.001

Notes. [a]Indicates the discrepancy of model 2 from model 1.
[b]Indicates the discrepancy of model 3 from model 2.
[c]Indicates the discrepancy of model 4 from model 1.
[d]Indicates the discrepancy of model 5 from model 4.

The initial Process-Product two-factor model was then modified such that expected correlations among items were allowed. Some redundancy was built into the RIBS, as is built into many tests, to minimize errors and increase internal consistency and inter-item reliability (Nunnally, 1978). Some items had similar wording (e.g., Items 9, 4 and 10 shared the same clause "have trouble..."). There was the possibility that participants responded to such wording instead of thinking of their ideational behaviours. Therefore, the two-factor model was modified to allow correlations among a few selected items – those that shared similar wording. Nine such correlations were allowed. As a result, goodness of fit indices significantly increased (see the modified two-factor model in Table 1).

The chi-squared deviance of the modified model was significantly smaller than the initial model. The discrepancy of chi-squared was significant, indicating the modified model was a significantly better model than the initial model. Also, the modified model showed CFI and TLI to be larger than .9, as well as RMSEA smaller than .05. (CFI = .930, TLI = .921, and RMSEA = .034). The correlation between the Process and Product scales of the RIBS was statistically significant (r = .77, p < .001). The factor loadings of a modified process and product model are presented in Table 5.

Discussion

These results suggest that of the four models tested the Product and Process model fit the data best. The Product and Process subscale scores were significantly correlated with one another. Although psychometric theory does not rely on tests of significance, nor shared variance, the correlation did imply that 59% of the variance was shared. This was to be expected, given that some of this may be method variance, and given that the entire RIBS does focus on ideation. In a sense it is reassuring that the single index scoring system was not well supported by the CFA. Recall here that this system was the weakest in terms of cognitive and creativity theory. Fluency has been used alone a number of times (e.g. Baer, 1991; Hocevar, 1979a, 1979b, 1980; Hong, Milgram & Gorsky, 1959; Milgram & Milgram, 1976) but only because there are often high bivariate correlations between fluency and originality, and between fluency and flexibility. Yet the one index approach is contrary to the standard definition, which is nearly universal in creativity research (Runco & Jaeger, 2012). This definition recognizes the necessary but not sufficient role of originality. It does not mention fluency; fluency is not a vital part of creativity. If one index of divergent thinking was to be used by itself, it should be originality not fluency. Originality is involved in all theories of creativity, but fluency is not.

Also, even though the correlation between fluency and originality tends to be quite high, implying a high level of overlap, fluency, originality and flexibility are far from redundant. More compellingly, there are studies suggesting that the unique variance of originality is reliable, even when the overlap and contribution from fluency is statistically removed (Runco, 1985; Runco & Albert 1985). The one index approach, with its complete reliance on fluency, does make research easier. The quality of ideas can be ignored, as are the complexities of scoring originality and flexibility. Scoring is simply a matter of counting ideas. But the one index approach was tested in the present research because it is a common practice, not because it is a good theoretical approach. Theories of creativity emphasize originality, not fluency, and theories of divergent thinking include several dimensions, not just one. The results of the CFA reported in this article confirm that it is better to use at least two indices when assessing ideation.

Study 1 suggested that the two scales of the RIBS representing ideational products and ideational processes best fit data from the children. Study 2 examined the two subscale model with data from the parents and teachers. The key question here concerns invariance. Does the model which fits the parents' responses best also fit the reponses from the teachers? The expectation was that there would be differences. This was based on previous comparisons of parents and teachers (e.g., Runco, Johnson & Bear, 1993).

STUDY 2

Method

Participants. A total of 796 parents and 90 teachers in the mid-Atlantic region of the United States participated in this study. The parents had children who participated in the CETA program. Several parents had multiple children in the CETA program; however their responses to the RIBS were made independently for each different child. Therefore, even if they participated in the survey multiple times, they were not counted as duplicates. The teachers completed the RIBS to describe each of their students.

Measure

The RIBS administered to the adults was as similar as possible to that given to the children, allowing for the fact that, unlike the children who completed the RIBS as a self-report, the adults were asked to rate children. This changed the instructions given on the RIBS such that they read:

"Use the 1-5 scale (given below) to indicate how often each of the phrases describes your child. Please indicate how you really see your child, not how you would prefer that he or she acts. You may need to approximate. Circle the response option that is THE CLOSEST to being accurate. [The word "student" was substituted for "child" in the teachers' version of the RIBS.] This survey contains 19 items which capture ideational behaviors of children. Responses are on a five-level Likert scale ranging from "Never (0)" through "Daily (4) (e.g., How often does your child/student come up with ideas you have never thought about?)."

Procedure

The parents received the RIBS in a packet that was sent home with their children. The teachers received the RIBS in packets dropped off at each participating school at designated data collection time points.

Results

Table 2 presents the means, standard deviations, skewness and kurtosis of the RIBS. In order to evaluate the violation of the nonnormality assumption, univariate skewness and univariate kurtosis were examined by cut-off scores. No values exceeded |2| for univariate skewness and |5| for univariate kurtosis, indicating the normality of distribution was supported.

Table 2
Descriptive Analysis of RIBS Items (Study 2)

Items	Factor	All (N=1333)			Parents (N = 693)			Teachers (N = 640)		
		M (SD)	Skewness	Kurtosis	M (SD)	Skewness	Kurtosis	M (SD)	Skewness	Kurtosis
Item1	Process	1.62 (1.34)	0.30	-1.13	1.87 (1.35)	0.10	-1.18	1.34 (1.28)	0.52	-0.93
Item2	Product	1.01 (1.02)	0.76	-0.20	1.12 (1.00)	0.67	-0.13	0.88 (1.02)	0.89	-0.16
Item5	Product	1.77 (1.18)	0.07	-0.89	1.90 (1.17)	-0.02	-0.81	1.63 (1.17)	0.16	-0.94
Item8	Product	2.02 (1.15)	-0.16	-0.73	2.54 (0.94)	-0.30	-0.12	1.46 (1.09)	0.25	-0.77
Item9	Process	2.24 (1.18)	-0.28	-0.72	2.64 (1.03)	-0.46	-0.30	1.80 (1.17)	0.01	-0.89
Item11	Product	2.04 (1.17)	-0.08	-0.76	2.54 (1.01)	-0.28	-0.42	1.50 (1.09)	0.25	-0.66
Item12	Product	1.75 (1.14)	0.15	-0.75	2.16 (1.05)	0.01	-0.60	1.32 (1.08)	0.42	-0.64
Item13	Process	2.12 (1.05)	-0.15	-0.51	2.51 (0.88)	-0.02	-0.47	1.70 (1.06)	0.10	-0.71
Item14	Product	1.75 (1.11)	0.06	-0.75	2.11 (1.01)	-0.13	-0.43	1.36 (1.08)	0.38	-0.65
Item16	Product	2.07 (1.10)	-0.23	-0.63	2.46 (0.93)	-0.18	-0.30	1.65 (1.12)	0.02	-0.99
Item19	Process	1.76 (1.06)	0.04	-0.57	2.06 (0.97)	-0.06	-0.22	1.43 (1.05)	0.28	-0.64
Item21	Product	1.56 (1.12)	0.33	-0.63	1.88 (1.08)	0.22	-0.60	1.21 (1.06)	0.54	-0.49
Item22	Product	2.19 (1.11)	-0.24	-0.58	2.26 (1.12)	-0.25	-0.64	2.11 (1.09)	-0.23	-0.51
Item23	Product	0.95 (1.07)	0.91	-0.10	1.15 (1.13)	0.70	-0.46	0.74 (0.96)	1.13	0.40
Item24	Product	1.61 (1.19)	0.23	-0.87	1.93 (1.12)	0.09	-0.75	1.26 (1.17)	0.50	-0.76
Item26	Process	1.46 (1.27)	0.48	-0.84	1.69 (1.28)	0.33	-0.94	1.21 (1.20)	0.64	-0.68
Item27	Product	1.83 (1.11)	0.04	-0.66	1.96 (1.09)	0.01	-0.60	1.70 (1.13)	0.10	-0.72
Item28	Product	1.60 (1.23)	0.23	-0.97	2.05 (1.20)	-0.17	-0.85	1.10 (1.07)	0.65	-0.41
Item29	Product	1.72 (1.21)	0.19	-0.90	2.20 (1.14)	-0.11	-0.75	1.21 (1.07)	0.54	-0.56

Note. M= Mean, SD = Standard Deviation

Tests of measurement invariance are useful for ensuring that observed differences between parents and teachers in the latent constructs are truly attributable to the true difference between them. Without establishing measurement invariance, the true difference in the underlying latent construct between parents and teachers could be confounded with measurement bias or different conceptualization, and therefore comparison between parents and teachers may be biased (Finney & Davis, 2003). The extent to which a confirmatory factor model measuring adults' perceptions of children's ideational behaviour exhibited measurement and structural invariance was determined using Mplus 7 (Muthén & Muthén, 1998-2012). Measurement invariance tests were conducted through a series of subsequent analyses by adding more invariance constraints to the sets of parameters be-

tween groups. The order was (a) allowing free all parameters, (b) constraining factor loadings to equal, and (c) constraining factor loadings and intercepts to equal. Robust maximum likelihood (MLR) estimation was employed for all analyses. A series of models tested were nested within each other. Therefore, the log likelihood ratio test was performed to compare nested models. Parents were used as the reference group in all invariance models.

In order to test measurement invariance between parents and teachers, both a single group analysis and a multi-group analysis were conducted (Byrne, 1998; Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). A baseline model was suggested by Study 1 (Chand O'Neal et al., 2014). This model had the construct of the creative Product and the construct of the creative Process.

Analyses with adult data indicated that the model representing these two subscales did not fit adult RIBS data well at all. This was not a huge surprise, given that (a) the earlier research had data from children and the data used here were from adults, and (b) the wording of the items in each subscale was slightly different (because of age and verbal differences in children and adults).

Quite possibly a Process-Product model could be improved by carefully examining the wording and correlations of the RIBS items for adults. With this in mind contraindicative items were eliminated from both the Product and Process scales. They showed low factor loadings on factors and contributed poor model fit. Additionally, all items were examined to ensure that they fit the overarching purpose of the RIBS, which was to represent the actual life experiences and opportunities for children's ideation. As a result, 16 items were retained for two scales (Item 1, 13, 26 were kept for Process, and item 2, 5, 11, 12, 14, 16, 21, 22, 23, 24, 27, 28, 29 were kept for Product).

Next inter-item consistencies and correlations among items were examined. Items were retained if they contributed to a higher reliability. (Items 9, 15, 19 were kept for Process and item 8 was kept for Product). In order to reduce redundancy of items, item 19 was retained but item 15, which was similar to item 19, was excluded. All of this led to significantly improved fit indices for the models. Importantly, this method of identifying items is similar to what has been used in previous research to improve model fits. DaVia Rubenstein, McCoach, and Siegle (2013), for example, also considered modification indices (i.e., expected chi-squared change by modifying models) offered by software in their item selection and refinement. They also examined item *Means and SDs*, keeping items with the lowest mean and larger SD, as well as sub-scale reliability to select items to be deleted for better model fit by checking reliability of items.

Invariance was then analyzed using these 19 items of the Product and Process scales. The list of items is provided in Table 2, which also offers information concerning excluded

items. Confirmatory factor analyses (CFAs) were separately conducted for each group to test the fit of the baseline models (see Table 3).

Table 3

Fit Indices for Confirmatory Factor Analysis of the Baseline Models

for Parents and Teachers

Model	CFI	TLI	RMSEA	SRMR	AIC	BIC	df	χ^2	Relative χ^2 (χ^2/df)
Parents	.906	.890	.056	.049	35263.00	35544.54	147	469.209	3.19
Teachers	.913	.899	.077	.047	29488.10	29764.71	147	699.594	4.76

Note. All ps < .001. CFI = comparative fit index; TLI = Tucker–Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Next, configural invariance was tested to determine if factor structures were equivalent across two groups. A configural invariance model was specified in which two-factor models were estimated simultaneously within each group. For the identification in each group, the factor mean was fixed at 0, and the first item's loading was fixed at 1 instead of fixing factor variances at 1. Because fixing factor variance at 1 implies constraining equal variances between two groups, the identification method of fixing factor variance at 1 was not appropriate. The baseline models of parents and teachers were compared by using multiple group CFA without imposing any equality constraints on any of the parameters. The result presented in Table 4 indicated that configural invariance was supported between parents and teachers, χ^2(294) = 1165.55, RMSEA = 0.067, CFI = .910, TLI=.896. Also, the models of two groups had the same number of factors, process and product, and the factor loading patterns were the same across groups: all factor loadings were significantly different from zero and had the same directional signs across groups (see Table 4).

Table 4

Fit Indices for Testing Measurement Invariance Models Between Parents and Teachers

Model	CFI	TLI	RMSEA	SRMR	AIC	BIC	df	χ^2	Relative χ^2 (χ^2/df)	$\Delta\chi^2$	Δdf
Configural Invariance Model	.910	.896	.067	.048	64751.1	65395.3	294	1165.55	3.96	_	_
Metric Invariance Model	.907	.898	.066	.053	64759.8	65315.7	311	1214.50	3.91	48.95[*]	17

Note.[*]p < .001

Hence, a series of model constraints was then applied in consecutive models to investigate potential declines of model fits resulting from measurement or structural non-invariance.

Lastly, the metric invariance model was tested to examine the equality of the unstandardized item factor loadings across groups by constraining factor loadings equal for both groups. The metric invariance model fit was good (see Table 4) but resulted in a significant decline in goodness-of-fit compared to the configural model, $\Delta\chi^2 = 48.95$, $\Delta df = 17$. This suggested that the factor loadings of parents were significantly different from teachers. The fact that metric invariance was rejected indicated that there was not even weak invariance, which in turn indicated that the items were related to the latent factors dissimilarly across groups. Put simply, the factor structures between the two groups remained the same, whereas items were weighted differently in both groups. Because a series of tests showed that measurement invariance was not met at the step of testing metric invariances, the subsequent measurement invariance tests were not conducted.

Table 5

Loading Pattern of Items on Latent Factors for a Modified Process-Product Model

Item	Process	Product
Item 4	0.498	
Item 5	-0.100	
Item 9	0.539	
Item 13	0.368	
Item 14	0.571	
Item 15	0.445	
Item 1		0.458
Item 2		0.514
Item 3		0.384
Item 6		0.493
Item 7		0.564
Item 8		0.521
Item 10		-0.084
Item 11		0.453
Item 12		0.486
Item 16		0.493
Item 17		0.645
Item 18		0.517
Item 19		0.613
Item 20		0.500
Item 21		0.486
Item 22		0.592
Item 23		0.493
Item 24		0.478
Item 25		0.637

Discussion

In general, the two factor model was more stable than the others. It was also better when contraindicative items were excluded. Study 2 relied on the two factor model without contraindicative items.

Analyses in Study 2, using the two factors without contraindicative items, did not support factorial invariance. The configural models looked similar between parents and teachers, with the number of factors and the directional sign of factor loadings remaining equal, but factor loadings were different for the two groups. Apparently the parents and teachers weighed RIBS items or perceived children's ideational behaviours in different ways.

This is not really a surprise, given previous empirical differences between parents and teachers. Runco (1987), for example, found that parents tended to weigh the cognitive aspects of creativity more heavily than teachers, and teachers weighed social aspects of creativity more heavily than parents. As was the case in Study 1, there was a notable correlation among factors. This too is not surprising. The RIBS was developed as a global ideational measure (Runco et al., 2000-2001). Subscales within the RIBS were hypothesized later, after the RIBS had demonstrated its usefulness (Chand O'Neal et al., 2014; Plucker et al., 2002; Runco et al., 2013). Recall here that Chand O'Neal et al. found the Product subscale to be particularly useful for distinguishing between students who had or had not participated in an extensive arts-integrated educational programme.

STUDY 3

The focus of Study 3 was on the degree of agreement between parents and teachers in terms of how strongly their RIBS ratings agreed with the ideational self-assessments of the children. This was determined by calculating product moment correlations. There is no need to reiterate the specifics of the Participants, Materials and Procedure. Each was identical with that reported for Studies 1-2.

Results and Discussion

Cronbach's alphas for the Product scale were slightly higher (.87, .85, and .93, for children, parents, and teachers, respectively) than were the alphas for the Process scale (.61, .67, and .82). Still, these each indicate that the RIBS scales had adequate reliability.

Product moment correlations indicated that Parents' RIBS Product scale ratings were significantly related to children's RIBS Product scale ratings ($r = .27$, $p < .01$). Ratings from Teachers on the RIBS Product scale were lower but still significantly related to children's Product scale ratings ($r = .19$, $p < .01$). Similarly, Parents' RIBS Process scale ratings were significantly correlated with Children's RIBS Process ratings ($r = .22$, $p < .01$), and again, Teachers' ratings were significantly related to Children's ratings on the Pro-

cess scale, though the coefficients were lower (r = .13, p < .01). When a total RIBS score (all items, regardless of Product or Process) was used, the Parent-Child correlation was .29 (p < .001) and the Teacher-Student correlation was .20 (p < .01).

These product moment coefficients were statistically significant but indicate that there was not much "shared variance" in the Parent-Child and Teacher-Student ratings. Still, this was not unexpected and not really a practical problem. It was not unexpected because adults and children hold different views and have different perspectives on creativity (Johnson, Runco & Raina, 2003; Miller & Sawyer, 1989; Noble, Runco & Ozkaragoz, 1993; Runco & Albert, 2005; Runco et al., 1993). The present findings do indicate that it would be unwise to generalize from one set of ratings (e.g., those given by parents) to either of the other groups (e.g., teachers or students themselves). Along the same lines it would be unwise to rely on one set of ratings, unless of course there is a reason to make decisions solely on the basis of one particular group. If research were to investigate the impact of creativity training on children's original thinking, and the hypothesis was that the training will be relevant to behaviour in the home, then parents' ratings might be the only or most important and ratings from the other groups could be ignored.

Table 6

Standardized Factor Loadings of the Baseline Model for the Whole Sample, Parents, and Teachers

Factor	Item	Whole Sample (N=1333)	Parents (N = 693)	Teachers (N = 640)
Process	Item1	0.51 (0.26)	0.34 (0.12)	0.52 (0.27)
	Item9	0.90 (0.81)	0.62 (0.38)	0.94 (0.89)
	Item13	0.87 (0.76)	0.60 (0.36)	0.92 (0.85)
	Item19	0.76 (0.58)	0.52 (0.27)	0.85 (0.73)
	Item26	0.59 (0.34)	0.51 (0.26)	0.57 (0.33)
Product	Item2	0.53 (0.29)	0.44 (0.19)	0.63 (0.40)
	Item5	0.62 (0.39)	0.46 (0.21)	0.77 (0.59)
	Item8	0.88 (0.78)	0.59 (0.35)	0.82 (0.68)
	Item11	0.85 (0.73)	0.56 (0.31)	0.81 (0.66)
	Item12	0.79 (0.63)	0.56 (0.32)	0.79 (0.62)
	Item14	0.82 (0.67)	0.63 (0.40)	0.82 (0.67)
	Item16	0.76 (0.57)	0.52 (0.27)	0.75 (0.57)
	Item21	0.80 (0.64)	0.65 (0.42)	0.82 (0.68)
	Item22	0.58 (0.34)	0.66 (0.43)	0.61 (0.38)
	Item23	0.56 (0.32)	0.46 (0.21)	0.61 (0.37)
	Item24	0.80 (0.63)	0.65 (0.43)	0.81 (0.65)
	Item27	0.72 (0.51)	0.70 (0.49)	0.81 (0.65)
	Item28	0.76 (0.57)	0.55 (0.30)	0.68 (0.47)
	Item29	0.79 (0.62)	0.57 (0.32)	0.71 (0.51)

Note. R^2 values are presented in parenthesis. All standardized factor loadings were statistically significant at p < .01.

General Discussion

Tests of ideation are probably the most commonly used measures of the potential for creativity thinking and problem solving. There is no such thing as a test of creativity, but tests can estimate the potential for certain expressions of creative talent. Every test is a sample of behaviour which estimates, more or less accurately, later performance. Every sample is smaller than the entire universe of behaviours being predicted, and every test suffers from some degree of measurement error.

Given the present results it might be tempting to rely on the Process and Product scales when using the RIBS. This approach is theoretically justified by Rhodes (1961) and Runco (2007) and by the results of Studies 1-2. Still, it is possible that research or a training programme hypothesizes something about or targets ideation that is independent of Process or Product. It might be wise to retain all RIBS items and choose scales based on the specifics of the research or training programme.

One point to keep in mind is that the RIBS was not developed to be a predictor. It was conceived as a criterion. Admittedly, a correlation can be done with two variables without insisting that one is a cause and one an effect (or predictor and criterion). Also, the RIBS assesses ideas, and ideas probably do contribute to various creative activities. This implies that the RIBS could be used as a predictor. Still, Wallach (1970) was quite adamant that DT should not be a criterion, and the RIBS is like DT in that the focus is ideation. Perhaps the question is, what is the primary concern when testing creative potential? If it is creativity, DT and the RIBS should not be criteria; they should only be predictors. If the concern is ideation, on the other hand, which is an important topic of study in its own right (Runco, 1999, 2013), then DT and the RIBS could be used as criteria.

The differences among parents and teachers uncovered in this research was not surprising, given earlier investigations showing that parents and teachers have divergent perspectives on children's creativity (Runco et al., 1991). Certainly parents and teachers are both likely to think about supporting children's creativity, but they experience children in such different contexts that they are likely to think differently about what children should and can do. Parents are likely to have in-depth experience with only a few children, usually their own offspring, while teachers will often have very broad experience with a large number of different children, but only in the educational setting. Educators may also think more about academic performances than parents, who might be as concerned about social and emotional issues as they are academic growth. The point is simply that it is no surprise that parents and teachers differ in their views of children's ideational behaviour.

The RIBS seems to be a reliable instrument. Apparently it can be used in different ways, but the results of the present investigation indicate that scoring the RIBS for Process and Product subscales does offer an interpretable approach. Future research might collect data from actual products and determine if the RIBS Product scale is in fact more strongly related to actual products than is the Process scale. There are good measures of creative products (e.g. Holland, 1961; Runco, 1987; Wallach & Wing, 1969), so this line of work would be easily done, as well as informative.

REFERENCES

Abedi, J. (2002). A latent-variable modeling approach to assessing reliability and validity of a creativity instrument. *Creativity Research Journal, 14*, 267-276.

Akaike, H. (1974). A new look at the statistical model identification. *IEE Transactions on Automatic Control, 19*, 716-23.

Baer, J. (1991). Generality of creativity across performance domains. *Creativity Research Journal, 4,* 23-39.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-46.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Chand O'Neal, I., Schulz Begle, A., Runco, M. A. (2014, August). *The Effects of Arts Integrated Instruction on Ideational Behavior in 4th and 5th Grade Students.* Presentation at the 122nd Annual Convention of the American Psychological Association, Washington, DC.

Chen, N., Roth, R. K. & Todhunter, J. E. (in press). I can do it! The effect of belief in stable luck on employee creativity. *Business Creativity and the Creative Economy*.

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

Cropley, D. & Cropley, A. (2012). A psychological taxonomy of organizational innovation: Resolving the paradoxes. *Creativity Research Journal, 24*, 29-40.

DaVia Rubenstein, L., McCoach, D. & Siegle, D. (2013). Teaching for Creativity Scales: An instrument to examine teachers' perceptions of factors that allow for the teaching of creativity. *Creativity Research Journal, 25*, 324-334.

Finney, S. J. & Davis, S. L. (2003, April). *Examining the invariance of the achievement goal questionnaire across gender.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*, 444-454.

Guilford, J. P. (1968). *Creativity, intelligence and their educational implications.* San Diego, CA: Knapp.

Hocevar, D. (1979a). A comparison of statistical infrequency and subjective judgment as criteria in the measurement of originality. *Journal of Personality Assessment, 43,* 297-299.

Hocevar, D. (1979b). Ideational fluency as a confounding factor in the measurement of originality. *Journal of Educational Psychology, 71*, 191-196.

Hocevar, D. (1980). Intelligence, divergent thinking, and creativity. *Intelligence, 4,* 25-40.

Holland, J. L. (1961). Creative and academic performance among talented adolescents. *Journal of Educational Psychology, 53,* 132-143.

Hong, E., Milgram, R. M. & Gorsky, H. (1959). Original Thinking as a Predictor of Creative Performance in Young Children. *Roeper Review, 18*, 147-49

Johnson, D., Runco, M. A. & Raina, M. K. (2003). Parents and teachers' implicit theories of children's creativity: A cross-cultural perspective. *Creativity Research Journal, 14*, 427-438.

MacCallum, R. C., Browne, M. W. & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130-149.

Milgram, R. M. & Milgram, N. A. (1976). Creative thinking and creative performance in Israeli students. *Journal of Educational Psychology, 68*, 255-259.

Miller, H. B. & Sawyers, J. K. (1989). A comparison of self and teachers ratings of creativity in fifth grade children. *Creative Child and Adult Quarterly, 14*, 179-185.

Muthén, L. K. & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.

Noble, E. P., Runco, M. A. & Ozkaragoz, T. Z. (1993). Creativity in alcoholic and nonalcoholic families. *Alcohol, 10*, 317-322.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

Plucker, J. A., Runco, M. A. & Woong, L. (2006). Predicting ideational behavior from divergent thinking and discretionary time on task. *Creativity Research Journal, 18*, 55-63.

Rhodes, M. (1961). An analysis of creativity. *Phi Delta Kappan, 42,* 305-310.

Richards, R. (1999). Everyday creativity. In: M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity*(1st ed.) (pp. 683-687). San Diego: Academic Press.

Runco, M. A. (1985). Reliability and convergent validity of ideational flexibility as a function of academic achievement. *Perceptual and Motor Skills, 61,* 1075-1081.

Runco, M. A. (1987). Interrater agreement on a socially valid measure of students' creativity. *Psychological Reports, 61,* 1009-1010.

Runco, M. A. (1999). *Divergent and creative thinking.* Cresskill, NJ: Hampton Press.

Runco, M. A. (2007). *Creativity. Theories and themes: Research, development, and practice.* Burlington, MA: Elsevier Academic Press.

Runco, M. A. (2008). Commentary: Divergent thinking is not synonymous with creativity. *Psychology of Aesthetics, Creativity, and the Arts, 2,* 93-96.

Runco, M. A. (2013, October). *The real creativity crisis.* Keynote address to the Mississippi Association for Gifted Children, Hattiesburg, MS.

Runco, M. A. & Albert, R. S. (1985). The originality and convergent validity of ideational originality in academically gifted and nongifted children. *Educational and Psychological Measurement, 45*, 483-501.

Runco, M. A. & Albert, R. S. (2005). Parents' personality and the creative potential of exceptionally gifted boys. *Creativity Research Journal, 17*, 355-368.

Runco, M. A. & Charles, R. (1993). Judgments of originality and appropriateness as predictors of creativity. *Personality and Individual Differences, 15*, 537-546.

Runco, M. A. & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24,* 92-96

Runco, M. A. & Plucker, J. A. & Lim, W. (2000-2001). Development and psychometric integrity of a measure of ideational behavior. *Creativity Research Journal, 13*, 393-400.

Runco, M. A., Johnson, D. & Bear, P. (1993). Parents' and teachers' implicit theories of children's creativity. *Child Study Journal, 23*, 91-113.

Runco, M. A., Okuda, S. M. & Thurston, B. J. (1991). Environmental cues and divergent thinking. In M. A. Runco (Ed.), *Divergent thinking* (pp. 79-85). Norwood, NJ: Ablex.

Runco, M. A., Walczyk, J. J., Acar, S., Cowger, E. L., Simundson, M. & Tripp, S. (2013). The incremental validity of a Short Form of the Ideational Behavior Scale and usefulness of distractor, contraindicative, and lie scales. *The Journal of Creative Behavior, 48*, 185-197.

Runco, M. & Richards, R. (1998). *Eminent creativity, everyday creativity, and health.* Stamford, CT: Ablex.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

Shapiro, A. K. (1960). A contribution to a history of the placebo effect. *Behavioral Science, 5*, 109-135.

Steenkamp, J. E. M. & Baumgartner, H. (1998). Assessing measurement invariance in crossnational consumer research. *Journal of Consumer Research, 25*, 78-90.

Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking: Norms – Technical manual.* Bensenville, IL: Scholastic Testing Service.

Torrance, E. P. (1995). *Why fly?* Cresskill, NJ: Hampton Press.

Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 4–70.

Vartanian, O., Bouak, F., Caldwell, J. L., Cheung, B., Cupchik, G., Jobidon, M. E., Lam, Q., Nakashima, A., Paul, M., Peng, H., Silvia, P. J. & Smith, I. (2014). The effects of a single night of sleep deprivation on fluency and prefrontal cortex function during divergent thinking. *Frontiers in Human Neuroscience, 8,* 1-12.

Wallach, M. A. (1970). Creativity. In P. H. Mussen (ed.), *Manual of child psychology* (Vol. 1, pp. 1211-1272). New York: Wiley.

Wallach, M. A. & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction.* New York, NY: Holt, Rinehart & Winston.

Wallach, M. A. & Wing, C. (1969). *The talented student.* New York, NY: Holt, Rinehart & Winston.

Corresponding author at: Mark A. Runco, Torrance Creativity Center, University of Georgia, Athens, 323 Aderhold Hall, Athens, GA 30602
E-mail: runco@uga.edu

Corresponding author at: Ivonne Chand O'Neal, Creativity Testing Services, LLC, Washington, DC USA
E-mail: ichand@comcast.net

Corresponding author at: Sue Hyeon Paek, University of Georgia, Athens, GA USA
E-mail address: suehpaek@gmail.com