Old or New? Evaluating the Old/New Scoring Method for Divergent Thinking Tasks

By: Paul J. Silvia, Emily C. Nusbaum, and Roger E. Beaty

This is the peer reviewed version of the following article:

Silvia, P.J., Nusbaum, E.C. and Beaty, R.E. (2017). Old or New? Evaluating the Old/New Scoring Method for Divergent Thinking Tasks. *Journal of Creative Behavior*, 51(3): 216-224. https://doi.org/10.1002/jocb.101

which has been published in final form at https://doi.org/10.1002/jocb.101. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Abstract:

When people generate responses during a divergent thinking task, some responses are "old" (retrieved from memory) and some are "new" (generated on the spot). K.J. Gilhooly, E. Fioratou, S.H. Anthony, and V. Wynn (2007) suggested that old and new responses stem from different cognitive strategies and differ in key ways. The present research explored the old/new scoring method in a sample of 143 young adults. After completing unusual uses tasks, the participants classified each response as old or new. The creativity of each response was also rated by three judges and by the participants themselves. As in past research, "old" responses appeared significantly earlier in the task and were rated as significantly less creative by both the judges and the participants. Old and new responses, however, correlated equally strongly with predictors of creative ability, such as openness to experience and its facets. Overall, the old/new scoring approach appears promising as a way of illuminating the diverse mental strategies people use to generate ideas.

Keywords: creativity | divergent thinking | assessment | personality | openness to experience | HEXACO

Article:

The basic notion behind divergent thinking tasks—to study how people come up with creative ideas, give them a simple prompt and ask them to come up with ideas—remains compelling after many decades of use. But like all open-ended tasks, divergent thinking tasks pose a challenge to researchers: what do you do with the huge heap of responses that people generate? Developing ways of quantifying these qualitative responses has been a major part of divergent thinking research. Recent years have seen a renewal of interest in assessment and scoring methods, from variations on classical frequency metrics (e.g., Plucker, Qian, & Wang, 2011) to subjective ratings (e.g., Benedek, Mühlmann, Jauk, & Neubauer, 2013; Silvia et al., 2008) to automated

scoring methods (e.g., Acar & Runco, 2014; Beaty, Silvia, Nusbaum, Jauk, & Benedek, 2014; Prabhakaran, Green, & Gray, 2014). In the present research, we evaluate the *old/new scoring method*, developed by Gilhooly, Fioratou, Anthony, and Wynn (2007), which has shown promise in their original study and in a recent fMRI study (Benedek et al., 2014a).

One appeal of the old/new scoring approach is its fiendish simplicity. First, people generate responses to a divergent thinking task, according to standard administration methods. After the task, people read their responses and classify each one as *old* or *new*. Old responses are ideas that people had seen or heard of before. As they were already stored in memory, these were ideas that predated the task and that were simply retrieved from memory. For example, common responses during a brick-use task are "to break a window" and "to keep a car from rolling backward." People do use bricks for these purposes, so for many respondents, the uses are remembered rather than generated. New responses, in contrast, are ideas that people generated on the spot. As a result, these responses are more properly thought of as *creative* because people came up with the idea by combining or transforming knowledge.

The old/new scoring method grew out of Gilhooly et al.'s (2007) analysis of people's strategies for unusual uses tasks. They found that most people started with a memory retrieval strategy: they searched memory for interesting and quirky examples of things one could use an object for. These responses were thus mostly "old," because they were uses people had encountered before and stored in memory. Over time, people tended to switch from trying to recall uses to applying generation strategies, such as disassembling an object or finding broad categories of uses (e.g., cooking, fighting) that an object could be repurposed for. These strategies are more executively demanding than memory retrieval, but they were more likely to result in "new" uses.

In the initial study, some differences between old and new responses were explored by asking a sample of 103 people to generate unusual uses for a tire or barrel and then classify each response as old or new (Gilhooly et al., 2007, Study 2). A single judge rated the creative quality of subsamples of 20 old and 20 new responses, and the new responses received higher novelty scores. During the task, old responses were more likely to appear early on and new responses were more likely to appear later, consistent with the idea that people tend to start with a simple memory retrieval strategy and then later shift to more complex strategies (Beaty & Silvia, 2012). Some evidence for the differential meaning of old and new responses came from their relations with verbal fluency tasks. The number of new uses was more strongly linked to letter fluency (a relatively more executively driven task), whereas the number of old uses was more strongly linked to category fluency (a relatively less executively driven task). In short, old and new responses do seem to stem from different strategies, and new responses appear to be more creative.

Some additional evidence for old/new scoring comes from a recent neuroimaging study. Benedek et al. (2014a) recruited 35 people to generate unusual objects in an MRI scanner. People had 60 seconds to generate uses for each object. After, people classified their own responses as old or new. Three raters then subjectively scored the responses. Many of Gilhooly et al.'s (2007) findings were replicated. The new responses were rated as more creative than the old ones, and people were more likely to give old responses at the start of the task. The neuroimaging results revealed that new responses, relative to old, were associated with greater

activity in the anterior part of the left inferior parietal cortex, a region involved in mental simulation (Schacter, Addis, & Buckner, 2007).

In the present research, we sought to evaluate old/new scoring more closely. Thus far, past work suggests that the method is promising, and a few key effects—new responses come later and are more creative—have been replicated. Not much evidence, however, addresses whether new responses show stronger relationships with markers of creativity, such as personality traits, cognitive abilities, and creative backgrounds. Gilhooly et al. (2007) found that old and new responses had different correlations with verbal fluency tasks; Benedek et al.'s (2014a) study found that new and old responses had some different brain activation patterns. Our study thus expanded the evidence for validity by assessing a broad range of factors relevant to creativity in a large and diverse sample. We sought to replicate some of the features of old/new scoring—such as whether new responses are more creative and appear later—as well as extend past research. In particular, we wanted to see if old and new responses showed differential relationships with established markers of creativity. For example, does openness to experience, a trait central to creativity, more strongly predict the creativity of new responses? If so, then researchers could likely find stronger effects by focusing on the subset of responses that best reflect the operation of creative strategies. The participants thus completed a pair of divergent thinking tasks along with a broad set of measures of personality and creative activities.

Method

Participants

A total of 151 students at the University of North Carolina at Greensboro (UNCG) participated as part of a voluntary research option in a psychology class. Eight participants were excluded because they were non-native speakers of English or because they had scores higher than 2 on an infrequency scale (an updated version of the Chapman and Chapman 1983 scale) intended to catch inattentive or random responding (Maniaci & Rogge, 2014; McKibben & Silvia, 2015), leaving a final sample of 143. Most people in the final sample were female (69%) and young (M = 19.22, SD = 3.06, Mdn = 19, Min/Max = 18, 51). The sample was racially and ethnically diverse, according to self-reports: 36% African American, 9% Asian American or Pacific Islander, 52% European American, 7% Hispanic or Latino/a, and 5% Native American. (People could select more than one option or decline to select any.)

Procedure

The research was approved by the UNCG Institutional Review Board. People took part in sessions of up to eight people. After providing informed consent, people learned that the study was about personality and creative thinking styles. The experimenter explained the divergent thinking tasks and gave instructions for the old/new judgments people would make. The study started with two divergent thinking tasks, followed by measures of personality and creative achievement.

Divergent thinking tasks

The participants completed two divergent thinking tasks: unusual uses for a *box* and a *rope*. As in our past research, we emphasized that people should try to "be creative" (Nusbaum, Silvia, & Beaty, 2014) and that the creativity of their ideas was more important than the quantity.[1] People had 3 minutes to generate as many ideas as they wished.

Self-rated creativity

After people generated their responses, the software presented their responses to them and asked for their self-ratings of creativity. The instructions said "Now we'd like to know your own opinions about how creative your ideas were. In the next part, each of your ideas will appear on the screen. Please rate how creative, in your own opinion, each idea is." The ideas were presented in the order in which they were generated, and people were asked to rate each idea (e.g., "How creative, in your opinion, was this use for a BOX?") on a 5-point scale (1 = *not at all creative*, 5 = *very creative*).

Old/new judgments

After rating each of their ideas, people gave old/new judgments for each one. The experimenter had described these judgments prior to the tasks, and the software reiterated the instructions. For example, for the box task people were told "For this part, you'll see each of your uses for a BOX again. Please rate whether each idea is OLD or NEW. OLD IDEAS are uses that you saw or heard before sometime in the past. These are uses that you remembered from somewhere. NEW IDEAS are uses that you came up with yourself during the study. You haven't seen or heard these before." People were presented each idea in the order in which it was generated, and they simply indicated whether it was "OLD: I saw or heard of this somewhere before" (scored 0) or "NEW: I came up with this on the spot during the experiment" (scored 1).

Subjective scoring of DT responses

The responses people gave to the box and rope task were scored for creativity by three raters (the three authors), all of whom had extensive experience with subjective scoring. As in our past work, we first deleted all information linking a response to a participant's other data or responses and all information about the response's serial order. The responses were then sorted alphabetically and scored independently using a 5-point scale (1 = *not at all creative*, 5 = *very creative*). Responses receiving the lowest scores tend to be actual uses for objects (e.g., using a box to contain or store common objects, using a rope to tie or fasten something) or highly common uses given by large proportions of the sample. Responses receiving higher scores tended to be quirky, clever, elaborate, and, for lack of a better word, creative.

---

[1] The experimenter emphasized that people should try to generate creative ideas at the start of the session, and the task's instructions reiterated this point. For example, the written instructions people read before the box task said:

> For this task, you'll be asked to come up with as many original and creative uses for a BOX as you can. The goal is to come up with *creative ideas*, which are ideas that strike people as clever, unusual, interesting, uncommon, humorous, innovative, or different. Your ideas do not have to be practical or realistic; they can be quirky, silly, or even strange, so long as they are CREATIVE rather than ordinary. You can enter as many ideas as you like; just press ENTER after each one. The task will take 3 minutes. You can type in as many ideas as you like until then, but *creative quality is more important than quantity*. It is better to have a few really good ideas than a lot of uncreative ones.

Personality

We measured personality with the 100-item HEXACO (Lee & Ashton, 2004), which measures the six factors in the HEXACO model (Ashton & Lee, 2007): Honesty–Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience. People responded to each item using a 5-point scale (1 = *strongly disagree*, 5 = *strongly agree*). Most of the factors resemble their counterparts in popular Big Five models, but Honesty–Humility and Agreeableness differ most (Lee & Ashton, 2012). We were primarily interested in openness to experience, which looms large in creativity research (Fürst, Ghisletta, & Lubart, 2015; Karwowski, Lebuda, Wisniewska, & Gralewski, 2013; Kaufman, 2013). The HEXACO provides a global openness to experience scale as well as subscales for four facets: *aesthetic appreciation*, *unconventionality*, *inquisitiveness*, and *creativity*.

To expand on the assessment of openness to experience, we included the O/I subscales from the Big Five Aspect Scales (BFAS; DeYoung, Quilty, & Peterson, 2007). The BFAS splits openness to experience into *openness* and *intellect* aspects, and this distinction has proven fertile in understanding how openness to experience influences creativity (Kaufman et al., 2015; Nusbaum & Silvia, 2011b). Each subscale has 10 items, and people respond on a 5-point scale (1 = *strongly disagree*, 5 = *strongly agree*).

Creative hobbies and achievements

We measured creative achievements with the Creative Achievement Questionnaire (CAQ; Carson, Peterson, & Higgins, 2005), a popular self-report scale that assesses accumulated creative achievements in 10 domains. Higher scores reflect increasing levels of observable, public accomplishments in a creative field. Common creative behaviors, akin to everyday creativity, were assessed with the Biographical Inventory of Creative Behaviors (BICB; Batey, 2007), a 34-item scale that lists common creative activities (e.g., "Composed a piece of music," "Written a short story") and asks people to check which ones they have done in the past year. Both the CAQ and BICB have good psychometric properties and are widely used in creativity research (for a review, see Silvia, Wigert, Reiter-Palmon, & Kaufman, 2012). We used a log-transformed average of the 10 domains as a global CAQ score.

Results

Data Reduction and Model Approach

For many analyses, the data have a nested, multilevel structure. People completed two tasks, so some of their scores—the old/new judgments, self-rated creativity, the raters' scores, and the responses' serial order—vary within person. Other variables vary between people, such as HEXACO, Openness/Intellect, BICB, and CAQ scores. Some of our research questions are within-person hypotheses, such as whether people's old/new judgments and the responses' serial order are associated with subjective ratings of creativity, so we used multilevel models to analyze the data for many of our hypotheses. The regression coefficients for the multilevel analyses are all unstandardized, given the complexity of establishing standardized solutions for

multilevel data (Snijders & Bosker, 1994). All models were estimated in Mplus 7.3 using maximum likelihood with robust standard errors.

Our main outcome was the rated creativity of the responses to the divergent thinking tasks. The raters' scores showed good internal consistency. Cronbach's alpha, computed from the estimated within-person correlations, was .65 for the box task and .67 for the rope task. The subjective ratings of the divergent thinking responses were modeled as a latent variable. At the within-person level, a latent "creativity" variable was formed using the three raters' scores as indicators. At the between-person level, a latent "creativity" variable was formed using the three rater's scores (technically the random intercepts that vary across within-level clusters). The between-level residual variances were fixed to zero because they are typically extremely small and convergence problems can arise when near-zero residuals are estimated. The factor loadings were constrained to be equal across levels to ensure that the variable was invariant across levels (Heck & Thomas, 2009). The intraclass correlation for this latent creativity variable was .34, so roughly a third of the variance in ratings was at the between-person level.

Features of Old and New Responses

How many?

We averaged the number of new, old, and total responses across the two divergent thinking tasks to create overall scores. On average, people generated 8.36 total responses per task (SD = 4.40, Min/Max = 1.50–26.00). Of these, an average of 3.62 were new (SD = 2.80, Min/Max = .00–20.50) and 4.75 were old (SD = 3.01, Min/Max = .50–5.00). The correlation between the number of old and new responses was small (r = .15 [.01, .28]).

Did the raters view the new ideas as more creative?

Were the new responses more creative than the old ones? People's old/new judgments were specified as the predictor (centered within person), and the latent creativity variable (described earlier) was the outcome. Old/new judgments significantly predicted creativity ratings, b = .37, SE = .03, p < .001. As in past research (Benedek et al., 2014a; Gilhooly et al., 2007), the new responses were rated as more creative than the old ones.

Did people see their new ideas as more creative?

Old/new judgments had the same relationship with people's own ratings of their responses. The participants' self-rating for each response was specified as the outcome, and old/new judgments were the within-person predictor (again centered within person). Old/new judgments significantly predicted self-rated creativity, b = 1.09, SE = .07, p < .001, so the participants rated their new responses as more creative than their old ones.

When did new ideas appear?

Finally, we explored whether the distribution of old and new responses changed across time. At the within-person level, we specified old/new judgments as the within-person predictor (centered

within person) and serial order—the serial position of a response, ranging from 1 to *k*—as the outcome. Old/new judgments significantly predicted serial order ($b = 1.22$, $SE = .26$, $p < .001$): new responses were more likely to be generated later, a finding that replicates past studies (Benedek et al., 2014a; Gilhooly et al., 2007).

As an aside, we also replicated the serial order effect for creativity ratings (Beaty & Silvia, 2012). When the raters' scores are the outcome, the serial order of the responses significantly predicted subjective ratings: later responses were rated as more creative ($b = .011$, $SE = .003$, $p < .001$).

In summary, old and new responses behaved as one would expect. "Old" responses retrieved from memory were more likely to appear early in the task, and "new" responses generated on the spot were more likely to come later, reflecting a shift from simple memory retrieval strategies to more complex ideation strategies (Gilhooly et al., 2007). "New" responses were also significantly more creative, as judged by independent raters and by the participants themselves.

Predictors of Response Quantity

Predictors of fluency—how many old, new, and total responses people generated—are shown in Table 1. These are conventional between-person models, so standardized effects are reported. For the total number of responses, notable effects appeared for the HEXACO Openness to Experience factor ($r = -.20$ [−.35, −.05]) and the BFAS Openness subscale ($r = -.24$ [−.38, −.11]). Both coefficients were negative, so people higher in openness generated fewer responses overall. Conscientiousness, in contrast, correlated positively with fluency ($r = .24$ [.07, .42]).

Table 1. Predictors of the Total Number of Responses and the Number of Old and New Responses

Predictor	Old	New	Total Number
Honesty–Humility	−.18 [−.36, .01]	−.06 [−.24, .11]	−.16 [−.36, .03]
Emotionality	−.01 [−.23, .21]	.08 [−.09, .25]	.04 [−.14, .23]
Extraversion	.07 [−.14, .28]	.15 [−.04, .35]	.14 [−.06, .35]
Agreeableness	−.21 [−.39, −.02]	.05 [−.11, .21]	−.11 [−.29, .07]
Conscientiousness	.24 [.04, .44]	.13 [−.04, .29]	.24 [.07, .42]
Openness to Experience	−.25 [−.41, −.09]	−.06 [−.20, .10]	−.20 [−.35, −.05]
O: Creativity	−.05 [−.22, .11]	.02 [−.13, .17]	−.03 [−.19, .14]
O: Unconventionality	−.15 [−.29, −.02]	−.01 [−.13, .12]	−.11 [−.23, .02]
O: Inquisitiveness	−.17 [−.32, −.03]	−.08 [−.25, .09]	−.17 [−.33, −.01]
O: Aesthetic Appreciation	−.20 [−.37, −.04]	−.06 [−.19, .08]	−.18 [−.33, −.03]
BFAS: Openness	−.27 [−.41, −.12]	−.10 [−.23, .04]	−.24 [−.38, −.11]
BFAS: Intellect	.01 [−.12, .14]	.01 [−.13, .15]	.01 [−.13, .15]
CAQ	−.01 [−.20, .18]	.03 [−.11, .17]	.01 [−.15, .17]
BICB	−.03 [−.18, .12]	.06 [−.15, .28]	.02 [−.16, .20]
Rated Creativity: Old	−.38 [−.51, −.24]	−.30 [−.43, −.17]	−.45 [−.58, −.31]
Rated Creativity: New	−.26 [−.41, −.10]	−.26 [−.38, −.13]	−.34 [−.48, −.19]

Note. The coefficients are correlations; 95% confidence intervals are in brackets. The HEXACO factors and the rated creativity factors were modeled as latent variables; all other predictors are observed variables. The scores for the box and rope tasks were averaged. CAQ = Creative Achievement Questionnaire; BICB = Biographical Inventory of Creative Behaviors.

Breaking down the responses into old and new responses (Table 1) shows that the effects were largely driven by the number of old responses people generated. Most of the correlations with the number of new responses were small in size. Openness to Experience ($r = -.25$ [−.41, −.09]), BFAS Openness ($r = -.27$ [−.41, −.12]), Agreeableness ($r = -.21$ [−.39, −.02]), and Conscientiousness ($r = .24$ [.04, .44]), however, significantly predicted the number of old responses. As a result, their effects on overall fluency were a result of their relationship with old responses: people higher in openness to experience (and several of its facets), openness, and agreeableness generated fewer old ideas, and people higher in conscientiousness generated more old ideas.

Predictors of Response Quality

We then turned to whether the creativity ratings of old and new responses differentially correlated with personality and creative backgrounds. For these between-person models, we specified latent old and new factors. As Figure 1 shows, each factor was defined by the creativity ratings for the three raters. Old and new responses received similar ratings: the correlation was $r = .75$ [.61, .89]. Table 1 shows the correlations between rated creativity and fluency; the correlations were all negative, so people who generated better ideas generated fewer ideas.[2] For example, quality and quantity correlated $r = -.38$ [−.51, −.24] for old ideas and $r = -.26$ [−.38, −.13] for new ideas.

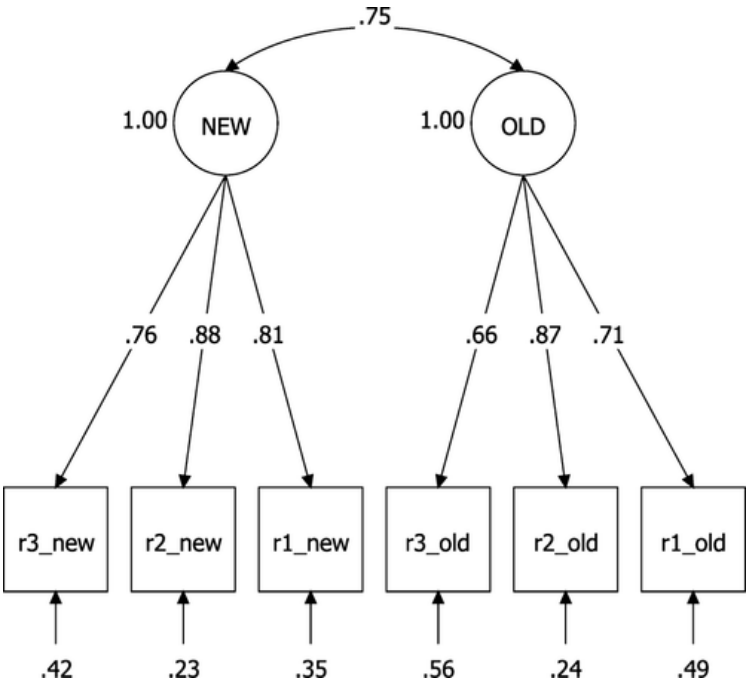

Figure 1. CFA and correlation between the rated creative quality of old and new responses.

[2] We commonly find small negative correlations between creative quality and quantity when we instruct people to "be creative" and use subjective scoring (Nusbaum et al., 2014). The small negative correlation is consistent with people adopting a strategic and controlled approach to divergent thinking (Nusbaum & Silvia, 2011a; Silvia, 2015). It is worth highlighting that traditional methods for divergent thinking—instructing people to generate as many ideas as possible and then coding for uniqueness—yields highly confounded measures of quality and quantity.

We next computed correlations between old and new creativity ratings (using the latent variables shown in Figure 1) and the HEXACO personality traits, which were specified as latent variables in which each factor was indicated by its four subscale scores (Lee & Ashton, 2004). Table 2 shows the effects. Several of the HEXACO factors—Honesty–Humility (largest $r = .14$), Emotionality (aka Neuroticism; largest $r = -.02$), and Agreeableness (largest $r = -.07$)—had at most small relationships with the rated creativity of the responses, consistent with most past research on personality and divergent thinking (e.g., Batey & Furnham, 2006). Conscientiousness had modest negative correlations with creativity, particularly for the new ($r = -.16$) responses over the old ($r = -.05$). Similarly, Extraversion had negative correlations with rated creativity—not a typical finding (Batey & Furnham, 2006)—and the effects were larger for the new ($r = -.20$) than for the old responses ($r = -.12$).

Table 2. Correlations with the Rated Creative Quality of Old and New Responses

Predictor	Old	New
Honesty–Humility	.14 [−.07, .35]	.11 [−.06, .29]
Emotionality	.01 [−.23, .24]	−.02 [−.21, .17]
Extraversion	−.12 [−.31, .08]	−.20 [−.42, .01]
Agreeableness	−.07 [−.27, .14]	.04 [−.19, .27]
Conscientiousness	−.05 [−.28, .18]	−.16 [−.37, .05]
Openness to Experience	.40 [.23, .57]	.30 [.11, .50]
O: Creativity	.22 [.05, .39]	.14 [−.03, .30]
O: Unconventionality	.23 [.05, .41]	.14 [−.01, .28]
O: Inquisitiveness	.09 [−.09, .27]	.05 [−.12,.22]
O: Aesthetic Appreciation	.38 [.24, .51]	.31 [.15, .47]
BFAS: Openness	.33 [.18, .48]	.12 [−.04, .29]
BFAS: Intellect	.09 [−.10, .29]	.08 [−.08, .24]
CAQ	.10 [−.07, .28]	.02 [−.14, .18]
BICB	−.03 [−.21, .16]	−.03 [−.19, .13]

Note. The coefficients are correlations; 95% confidence intervals are in brackets. The HEXACO factors and the rated creativity factors were modeled as latent variables; all other predictors are observed variables. The scores for the box and rope tasks were averaged. CAQ = Creative Achievement Questionnaire; BICB = Biographical Inventory of Creative Behaviors.

Openness to experience had the largest effect sizes of the HEXACO traits. The effects, however, were similar regardless of whether the new ($r = .30$) or old ($r = .40$) responses were analyzed. To explore openness to experience in more detail, we examined the correlations between its four facets and the creativity of the responses; the correlations are shown in Table 2. The correlations for the facets mirrored the effects for the higher-order factor: old and new responses generally had the same relationships.

We conducted similar analyses for the BFAS openness and intellect factors and for the CAQ and BICB, the two measures of creative activities and achievements. These four variables, treated as observed scores, were correlated with the latent creativity ratings for all responses and for the old and new subsets. Table 2 displays the correlations. The CAQ, BICB, and BFAS intellect scales had small effects at most (rs < .10). The BFAS openness scale had larger effect sizes, and its relationships with the creativity of the responses was somewhat lower for the new ($r = .12$) than the old ($r = .33$) responses.

Discussion

The present research examined old/new scoring, an intriguing approach to assessing divergent thinking that can give insight into how people generate their ideas. Many recent scoring methods suggest that divergent thinking tasks are more efficient and valid when subsets of responses are analyzed, such as the two responses people see as their best (Silvia et al., 2008) or the top two or three based on raters' scores (Benedek et al., 2013). Old/new scoring can be seen as an instance of a subset approach: old and new responses likely stem from different cognitive strategies, so they probably have different relationships with other markers of creativity.

As new scores more likely reflect creative processes, we wanted to evaluate the differential validity of the old and new responses. Our study replicated several findings that appeared in both past studies (Benedek et al., 2014a; Gilhooly et al., 2007). The raters judged the new responses as more creative than the old responses, and the new responses appeared later in the task, consistent with a shift from initially retrieving ideas to applying generative strategies. In addition, when asked to rate their own responses, the participants rated their new responses as more creative than their old ones.

Our findings do not fully support, however, the possibility of carving out the old from the new responses when assessing a person's divergent thinking ability. The creative quality of old and new responses were highly correlated, and people who received higher creativity scores from the raters—people high in openness to experience and BFAS openness—generally got higher scores on all their responses. The correlations between these traits were essentially the same for the old and the new responses (see Table 2), so the subset of new responses did not yield larger effect sizes. For quantity, however, some differences were found: people higher in openness to experience, for example, generated fewer old responses.

In addition, the pattern of fluency scores suggests that analyzing only the new responses might be impractical. People varied substantially in how many responses they generated, as in most divergent thinking studies, but they also varied in how many new responses they generated. Some people generated only a few new responses, and a few generated none. As a result, the data are sparse—or missing—for some participants if only the new responses are analyzed. Other subset methods hold the number constant across respondents, such as the best two or three responses, so the number of responses available to score—and hence the reliability of the scores—is not affected by a respondent's overall fluency.

The differential validity of old and new responses should be explored further in future work. Our study emphasized predictors related to personality traits and creative activities. We did not find notably different correlations for old and new responses, but it is likely that predictors related to cognitive abilities—such as fluid intelligence, verbal knowledge, and associative and retrieval abilities—would differentially predict old and new responses more sharply. A range of cognitive abilities foster better divergent thinking (Benedek, Jauk, Sommer, Arendasy, & Neubauer, 2014b; Benedek, Könen, & Neubauer, 2012; Silvia & Beaty, 2012; Silvia, Beaty, & Nusbaum, 2013) in ways that could differentiate old and new responses. For example, people higher in fluid intelligence (Gf) are more able to apply an ideational strategy that yields good

ideas (Nusbaum & Silvia, 2011a, Study 2). Likewise, people higher in broad retrieval ability can manage directed search processes in memory more effectively, making it easier to retrieve knowledge that is not highly accessible (Avitia & Kaufman, 2014; Silvia et al., 2013). It is possible that cognitive abilities, not personality traits, are where different relationships with old and new responses will be found.

But regardless of whether other factors differentially predict the creative quality of old and new responses, the old/new scoring method lends some useful insights into the process of divergent thinking. The method reinforces the notion that people call upon diverse strategies when trying to generate ideas, and having participants classify their responses as old or new would be useful for studies interested in cognitive processes involved in idea generation. Overall, the old/new method is straightforward, appears promising, and deserves more attention in future research.

References

Acar, S., & Runco, M.A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal*, **26**, 229– 238.

Ashton, M.C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, **11**, 150– 166.

Avitia, M.J., & Kaufman, J.C. (2014). Beyond *g* and *c*: The relationship of rated creativity to long-term storage and retrieval (*Glr*). *Psychology of Aesthetics, Creativity, and the Arts*, **8**, 293– 302.

Batey, M. (2007). *A psychometric investigation of everyday creativity*. Unpublished doctoral dissertation, University College, London.

Batey, M., & Furnham, A. (2006). Creativity, intelligence, and personality: A critical review of the scattered literature. *Genetic, Social, and General Psychology Monographs*, **132**, 355– 429.

Beaty, R.E., & Silvia, P.J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*, **6**, 309– 319.

Beaty, R.E., Silvia, P.J., Nusbaum, E.C., Jauk, E., & Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Memory and Cognition*, **42**, 1186– 1197.

Benedek, M., Jauk, E., Fink, A., Koschutnig, K., Reishofer, G., Ebner, F., & Neubauer, A.C. (2014a). To create or to recall? Neural mechanisms underlying the generation of creative new ideas. *NeuroImage*, **88**, 125– 133.

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A.C. (2014b). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence*, **46**, 73– 83.

Benedek, M., Könen, T., & Neubauer, A.C. (2012). Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts*, **6**, 273– 281.

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.

Carson, S.H., Peterson, J.B., & Higgins, D.M. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. *Creativity Research Journal*, **17**, 37– 50.

Chapman, L.J., & Chapman, J.P. (1983). *Infrequency Scale*. Unpublished test, T.R. Kwapil, Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC.

DeYoung, C.G., Quilty, L.C., & Peterson, J.B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, **93**, 880– 896.

Fürst, G., Ghisletta, P., & Lubart, T. (2015). Toward an integrative model of creativity and personality: Theoretical suggestions and preliminary empirical testing. *Journal of Creative Behavior*, DOI: 10.1002/jocb.71.

Gilhooly, K.J., Fioratou, E., Anthony, S.H., & Wynn, V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology*, **98**, 611– 625.

Heck, R.H., & Thomas, S.L. (2009). *An introduction to multilevel modeling techniques* (2nd edn). New York: Routledge.

Karwowski, M., Lebuda, I., Wisniewska, E., & Gralewski, J. (2013). Big Five personality traits as the predictors of creative self-efficacy and creative personal identity: Does gender matter? *Journal of Creative Behavior*, **47**, 215– 232.

Kaufman, S.B. (2013). Opening up openness to experience: A four-factor model and relations to creative achievement in the arts and sciences. *Journal of Creative Behavior*, **47**, 233– 255.

Kaufman, S.B., Quilty, L.C., Grazioplene, R.G., Hirsh, J.B., Gray, J.R., Peterson, J.B., & DeYoung, C.G. (2015). Openness to experience and intellect differentially predict creative achievement in the arts and sciences. *Journal of Personality*, DOI: 10.1111/jopy.12156.

Lee, K., & Ashton, M.C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, **39**, 329– 358.

Lee, K., & Ashton, M.C. (2012). *The H factor of personality: Why some people are manipulative, self-entitled, materialistic, and exploitive—and why it matters for everyone*. Waterloo, Canada: Wilfrid Laurier University Press.

Maniaci, M.R., & Rogge, R.D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, **48**, 61- 83.

McKibben, W.B., & Silvia, P.J. (2015). Evaluating the distorting effects of inattentive responding and social desirability on self-report scales in creativity and the arts. *Journal of Creative Behavior*, DOI: 10.1002/jocb.86.

Nusbaum, E.C., & Silvia, P.J. (2011a). Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence*, **39**, 36– 45.

Nusbaum, E.C., & Silvia, P.J. (2011b). Are openness and intellect distinct aspects of openness to experience? A test of the O/I model. *Personality and Individual Differences*, **51**, 571– 574.

Nusbaum, E.C., Silvia, P.J., & Beaty, R.E. (2014). Ready, set, create: What instructing people to "be creative" reveals about the meaning and mechanisms of divergent thinking. *Psychology of Aesthetics, Creativity, and the Arts*, **8**, 423– 432.

Plucker, J.A., Qian, M., & Wang, S. (2011). Is originality in the eye of the beholder? Comparison of scoring techniques in the assessment of divergent thinking. *Journal of Creative Behavior*, **45**, 1– 22.

Prabhakaran, R., Green, A.E., & Gray, J.R. (2014). Thin slices of creativity: Using single-word utterances to assess creative cognition. *Behavior Research Methods*, **46**, 641– 659.

Schacter, D.L., Addis, D.R., & Buckner, R.L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, **8**, 657– 661.

Silvia, P.J. (2015). Intelligence and creativity are pretty similar after all. *Educational Psychology Review*, DOI: 10.1007/s10648-015-9299-1.

Silvia, P.J., & Beaty, R.E. (2012). Making creative metaphors: The importance of fluid intelligence for creative thought. *Intelligence*, **40**, 343– 351.

Silvia, P.J., Beaty, R.E., & Nusbaum, E.C. (2013). Verbal fluency and creativity: General and specific contributions of broad retrieval ability (Gr) factors to divergent thinking. *Intelligence*, **41**, 328– 340.

Silvia, P.J., Wigert, B., Reiter-Palmon, R., & Kaufman, J.C. (2012). Assessing creativity with self-report scales: A review and empirical evaluation. *Psychology of Aesthetics, Creativity, and the Arts*, **6**, 1– 16.

Silvia, P.J., Winterstein, B.P., Willse, J.T., Barona, C.M., Cram, J.T., Hess, K.I., Martinez, J. L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, **2**, 68– 85.

Snijders, T.A.B., & Bosker, R.J. (1994). Modeled variance in two-level models. *Sociological Methods and Research*, **22**, 342– 363.