

In press, *Thinking Skills and Creativity*, February 2, 2021

Measuring Everyday Creativity:

A Rasch Model Analysis of the Biographical Inventory of Creative Behaviors (BICB) Scale

Paul J. Silvia, Rebekah M. Rodriguez

University of North Carolina at Greensboro

Roger E. Beaty, Emily Frith

Pennsylvania State University

James C. Kaufman

University of Connecticut

Paul Loprinzi

University of Mississippi

Roni Reiter-Palmon

University of Nebraska at Omaha

Author Note

Paul J. Silvia and Rebekah M. Rodriguez, Department of Psychology, University of North Carolina at Greensboro; Roger E. Beaty and Emily Frith, Department of Psychology, Pennsylvania State University; James C. Kaufman, Department of Educational Psychology, University of Connecticut; Paul Loprinzi, Department of Health, Exercise Science, and Recreation Management, University of Mississippi; Roni Reiter-Palmon, Department of Psychology, University of Nebraska at Omaha. The data and input files are available at Open Science Framework: <https://osf.io/dkrhy/>. Address correspondence to Paul J. Silvia, Department of Psychology, P.O. Box 26170, University of North Carolina at Greensboro, Greensboro, NC, 27402-6170, e-mail: p\_silvia@uncg.edu.

### Abstract

Research on everyday creativity—the “little c” creative activities people do in their everyday lives—commonly uses self-report scales to assess people’s engagement in different activities. The present research presents a detailed psychometric analysis of the Biographical Inventory of Creative Behaviors (BICB), a 34-item *yes/no* checklist of common creative activities that has become one of the most popular self-report measures of everyday creative behaviors. Based on a sample of 2,359 adults, the reliability, dimensionality, item fit, item difficulty, and test information were evaluated from a Rasch model perspective. Overall, the BICB shows good evidence for score reliability and appears essentially unidimensional; a small cluster of misfitting and locally dependent items were flagged for impairing unidimensionality. The items’ difficulty level was generally moderate and suitable for the scale’s intended populations and purposes. Differential item functioning (DIF) based on gender and age, estimated via Rasch tree recursive partitioning methods, found notable gender-based DIF (generally reflecting culturally gendered qualities of some creative activities) but little age-based DIF. Taken together, the BICB has many psychometric strengths. Some opportunities for future scale refinement are discussed.

**Keywords:** Biographical Inventory of Creative Behaviors; everyday creativity; creativity assessment; Rasch models

## 1. Introduction

Not all creativity is world-changing, “Big C” level genius. The tradition of creativity research interested in *everyday creativity* focuses on the diverse types of activity that people engage in during everyday life (Cotter et al., 2019; Fürst & Grin, 2018; Richards, 2010). These “mini-c” and “little-c” forms of creativity (Kaufman & Beghetto, 2009) make up the vast bulk of human creativity—there are many more drawings and paintings found on refrigerators than on gallery walls—and reveal much about the central role of creativity in motivation and well-being (Conner et al., 2018; Cotter et al. 2019; Richards, 2007).

The Biographical Inventory of Creative Behaviors (BICB), developed by Batey (2007), has emerged as one of the more popular self-report tools to measure everyday creativity. It is cited and discussed in many major reviews of self-report tools in creativity research (Kaufman, 2019; Puryear et al., 2017; Said-Metwaly et al., 2017; Silvia et al., 2012) and has been used in research on wide-ranging topics in creativity studies (e.g., Tempest & Radel, 2019; West & Somer, 2020). To date, however, the BICB has yet to receive a detailed psychometric evaluation. In the present research, we present a Rasch analysis of the BICB, with an emphasis on (1) the scale’s dimensionality and reliability, (2) the items’ difficulty and fit to the Rasch model, and (3) possible differential item functioning, evaluated using Rasch trees, based on gender and age. We conclude with an overall evaluation of the BICB’s psychometric properties.

### 1.1. Basics of the BICB

To measure people’s engagement in everyday creative behaviors, researchers have applied a few major assessment approaches. One approach uses experience sampling and diary methods (Silvia & Cotter, 2021) to track people’s activities during their typical days and weeks (Karwowski et al., 2017; Silvia et al., 2017). Although insightful, this approach is intricate and laborious, making it impractical for many research contexts (Cotter & Silvia, 2019). Another approach uses self-report assessments, such as rating scales and behavior checklists, to measure engagement in everyday creativity (e.g., Batey, 2007; Dollinger, 2003; Elisondo, 2020). These

approaches offer less detail and potentially less accuracy than daily-life methods but can be administered to large samples, so they afford valuable information about everyday creativity for a wide range of populations and research problems. In between these approaches are hybrids that combine self-report scales with other tools, such as performance tasks and peer reports (e.g., Fürst & Grin, 2018).

The BICB falls squarely within the self-report approach. It was developed and reported by Batey (2007) as part of a doctoral dissertation. Although the scale was never formally published, it proved useful in early research (Batey et al., 2010; Batey & Furnham, 2008; Furnham et al., 2008) and caught on quickly among creativity researchers, probably because it offers information about a broad range of activities in a compact, easy to administer scale. In a popular open-science archive of scales and research tools for research on creativity and the arts (<https://osf.io/4s9p6/>), the BICB is among the all-time most downloaded research tools.

The BICB consists of 34 items that describe common creative activities. The instructions ask participants to endorse “the activities you have been actively involved in” during the past year. For each item, people thus indicate if, in the past 12 months, they have “written a short story” (item 1) or “designed and planted a garden” (item 25). Table 1 lists abbreviated item stems. The items are diverse and wide-ranging, much more so than many self-report measures of creative activities. They include common activities related to the visual and performing arts and creative writing, intellectual and scientific activities, and interpersonal activities involving coaching, mentoring, and leadership. The BICB uses a binary checklist response scale, so people indicate simply if they did (*Yes* = 1) or did not (*No* = 0) actively engage in each activity during the past year. The scale is intended to yield a single score—usually a sum of the 34 items or an average of the 0/1 responses (i.e., the proportion of items endorsed)—so it has no subscales or facets. The brevity of the scale and the simplicity of the instructions and response format surely play a large part in the scale’s popularity among researchers.

The BICB can be contrasted with other popular self-report tools in creativity research.

First, its focus on common, everyday behaviors distinguishes the BICB from measures of *creative achievement*, which focus on major, public creative accomplishments that people have accumulated over time (e.g., Carson et al., 2005; Diedrich et al., 2018), such as awards, honors, and landmark creative works. Measures of *creative self-concepts*, as another point of contrast, focus on measuring people's beliefs about their creative traits and abilities, such as their views of their levels of creativity in different areas (McKay et al., 2017) or their self-efficacy for generating ideas (Karwowski et al., 2018, 2019).

Within the category of measures of everyday creativity, the BICB's focus on capturing people's engagement across a wide range of different behaviors distinguishes it from scales that assess motives for engaging in everyday creative actions (e.g., learning new things or coping with stress; Benedek et al., 2020). The BICB most resembles Dollinger's (2003) brief version of the Creative Behavior Inventory (CBI), which was created by selecting items from the larger scale first developed by Hocevar (1979). Like the BICB, the brief CBI asks for self-reported engagement in an array of everyday activities, but it has two key differences: (1) the CBI items focus on traditional arts and crafts domains, and (2) respondents rate how often they have done the activities to date, so the CBI measures cumulative creative activities. The BICB, in contrast, casts a wider net over creative activities and uses a time-window of 12 months, so it measures *recent* engagement in everyday creativity instead of *lifetime* engagement.

Research using the BICB has provided good evidence for validity. The BICB correlates positively with many other outcomes that a measure of everyday creativity should correlate with. People with high BICB scores, for example, also score higher on the Creative Achievement Questionnaire ( $r = .37$ ; Silvia et al., 2012), Creative Behavior Inventory ( $r = .53$ ; Silvia et al., 2012), self-rated creativity in different domains (the revised Creative Domains Questionnaire:  $r = .39$ , Silvia et al., 2012;  $r = .34$ , Wigert et al., 2012), divergent thinking fluency ( $r = .22$ , Furnham et al., 2008;  $r = .21$ , Batey et al., 2010), and openness to experience ( $r = .38$ , Furnham et al., 2008;  $r = .33$ , Batey et al., 2010).

## **1.2. The Present Research**

In the present research, we conducted a large-sample psychometric evaluation of the BICB. Given the scale's popularity, it's worth examining its strengths and weaknesses to provide scale users with practical knowledge about the scale's properties and to suggest some fruitful opportunities for future refinement and revision of the scale. Using a sample of over 2,300 adults, we conducted a Rasch analysis of the BICB with an eye toward key psychometric features: (1) the scale's dimensionality; (2) the items' difficulty and the scale's region of greatest reliability; (3) the possibility of item bias due to gender or age.

## **2. Method**

### **2.1. Participants**

The sample consisted of 2,359 adults who took part in one of a variety of studies that included the BICB. The data were pooled from many projects conducted over the past 10 years to yield a large sample. Of the total sample, 1,090 were participants enrolled at the University of Nebraska at Omaha and California State University, San Bernardino, whose responses were used in an earlier analysis of self-report measures of creativity (Silvia et al., 2012); 634 were students enrolled at the University of Mississippi who took part in a study of exercise and creativity (Frith & Loprinzi, 2020); and the remaining 635 participants were students at the University of North Carolina at Greensboro (UNCG) or community adults from the surrounding area who took part in one of many research projects on individual differences in creativity that included the BICB. All the projects had a primary focus on creativity except for a project focused on depression and motivation (Silvia et al., 2020). The samples had been screened for data quality, and there were no missing observations. The sample was predominantly female (1716 women, 643 men) and young ( $M = 22.20$ ,  $SD = 6.28$ ,  $Mdn = 20$ , range from 18 to 72 years old). The individual research projects did not specifically seek to oversample women, but it is common for research using American students recruited via psychology classes to have more women than men. This general trend is especially pronounced at UNCG, a former women's

college with a student population that is nearly 70% female.

## **2.2. Analysis Approach**

The analyses were conducted in R (R Core Team, 2020) using *psych* (Revelle, 2020), *psychotree* (Zeileis et al., 2020), and *TAM* (Robitzsch et al., 2020). The Rasch models were estimated in TAM using marginal maximum likelihood and identification via case constraint, which yields underlying trait scores centered on zero.

## **3. Results**

### **3.1. Rasch Model Fit**

The fit for the Rasch model was compared to fit for a 2PL IRT model, which estimates each's item's discrimination and adds 33 model parameters. Because of the large sample size, we compared the models using information theory criteria, such as the Gilula-Haberman log penalty (GHP), Akaike information criterion (AIC), and Bayesian information criterion (BIC), all of which penalize model complexity to varying degrees and indicate better fit with smaller values. Model fit was highly similar. The Rasch and 2PL models had nearly identical GHP values (.437 vs .436), the Rasch model had a slightly larger AIC than the 2PL (70,114.31 vs 69,924.62), and the BIC values were nearly identical but favored the Rasch model (70,316.12 vs 70,316.71). The high degree of similarity is unusual. Taken as a whole, the fit indices didn't clearly favor the more complex 2PL model. When an increase in model complexity is not apparently rewarded by improved model fit, it is reasonable to prefer the more parsimonious model (Bond et al., 2020), so we selected the Rasch model as our framework.

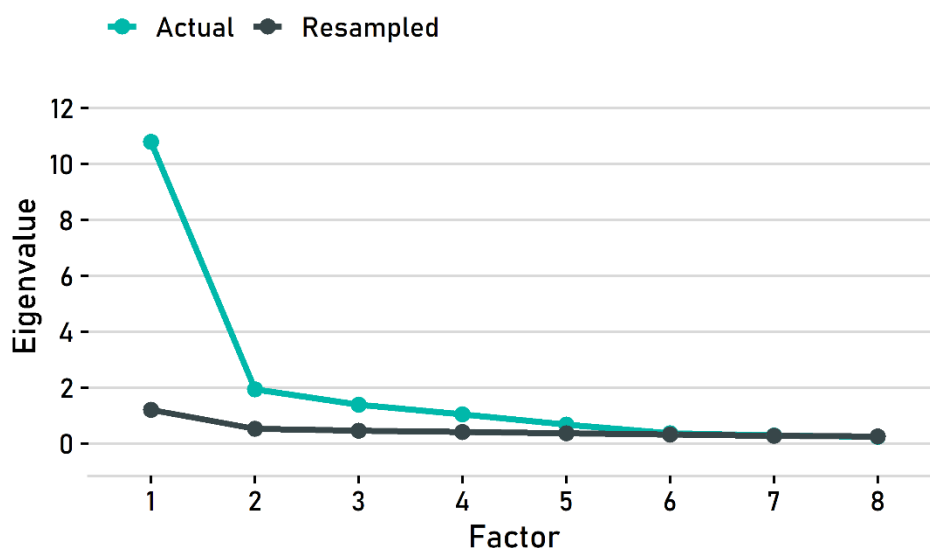
### **3.2. Reliability and Dimensionality**

Cronbach's alpha was high ( $\alpha = 0.86$ ), suggesting good internal consistency. Omega-total was very high ( $\omega_T = .95$ ). Omega-hierarchical, however, was much lower ( $\omega_H = .58$ ). Because  $\omega_H$  captures the degree to which the items are saturated by the general, common factor, it is worth closely evaluating the dimensionality of the BICB.

To explore dimensionality, we used several criteria to evaluate *essential*

*unidimensionality*, a less stringent criterion commonly applied to psychological constructs that recognizes that they are rarely strictly unidimensional even when the scores are dominated one factor (Slocum-Gori & Zumbo, 2011). We applied several methods: Horn's parallel analysis (Hayton et al., 2004), the ratio of the first-to-second eigenvalues (e.g., greater than 3:1 or 4:1; Slocum-Gori & Zumbo, 2011), and the minimum average partial (MAP) criterion (Velicer, 1976). The factor analyses were conducted in *psych* (Revelle, 2020) using maximum likelihood factor analysis. The correlations were modeled as tetrachoric because of the dichotomous response format.

**Figure 1.** Scree plot from a parallel factor analysis of the BICB items.



**Note.** For clarity, only the first 8 factors are depicted.

Overall, the evidence is consistent with essential (but not strict) unidimensionality. The MAP suggested 4 factors, and the parallel analysis suggested 6 factors, but the scree plots for the actual and resampled parallel analysis data showed a dominant first factor and only minor remaining factors (see Figure 1). The ratio of the first to second eigenvalues was 5.55:1, which is greater than conventional 3:1 and 4:1 guidelines and consistent with a dominant first factor



(Slocum-Gori & Zumbo, 2011).

We evaluated the meaning of the first factor versus the smaller factors using an exploratory factor analysis with a bifactor rotation, which estimates a common, general factor and then identifies specific, orthogonal factors (Jennrich & Bentler, 2011). The BICB loadings all loaded well on the general factor (loadings ranged from .35 to .73), with only one item (item 31) loading below .40. The specific factors did not consist of substantively meaningful facets but were locally dependent item pairs, which we examined in more detail.

A scale's unidimensionality can be eroded by *local dependence*, residual covariation between items remaining after accounting for the underlying latent trait (Chen & Thissen, 1997). We estimated it using the adjusted  $Q_3$  ( $aQ_3$ ) statistic, which corrects for the well-known negative bias in Yen's (1984)  $Q_3$  by centering the values on their mean (Marais, 2013). Flagging residual correlations over  $|.20|$  (Christensen et al., 2017) yielded 6 pairs of BICB items with notable local dependence. Local dependence can come from many sources, but in the BICB it largely reflected overlap in the creative activities:

- publishing an article and publishing research (items 11 and 23;  $aQ_3 = .31$ )
- being selected to lead or manage others and being made the leader of a group or team (19 and 32;  $aQ_3 = .27$ )
- critically evaluating a theory and producing a theory (items 13 and 17;  $aQ_3 = .26$ )
- drawing a cartoon and producing a picture (items 8 and 10;  $aQ_3 = .23$ )
- writing a novel and producing a script (items 2 and 4;  $aQ_3 = .21$ )
- producing a script and acting in a dramatic production (items 4 and 27;  $aQ_3 = .21$ ).

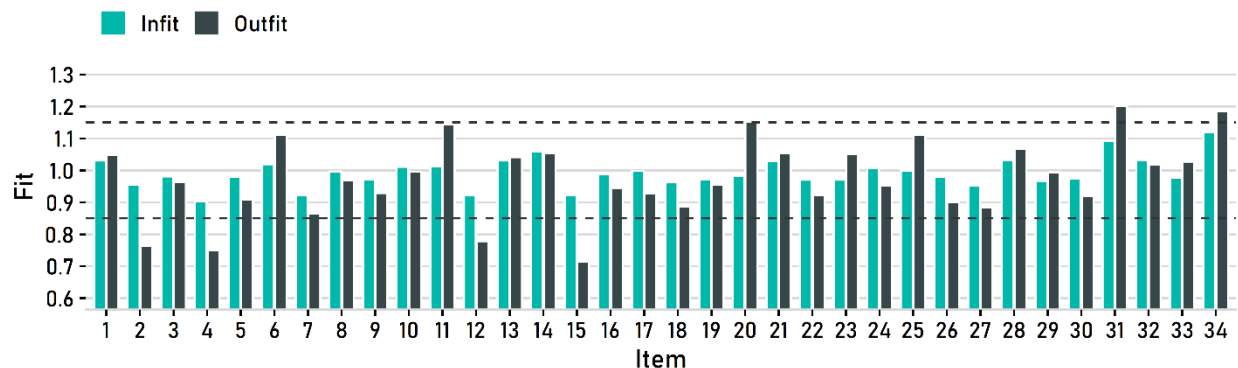
These local dependence statistics are useful because they highlight the low-hanging fruit for shortening the BICB. Most of these pairs represent relatively redundant items, usually with one being more general than another (e.g., producing a picture vs drawing a cartoon, a kind of picture). Trimming the relatively redundant items would abbreviate the scale while improving

its unidimensionality.

### 3.3. Item Fit

Item fit was evaluated with Infit and Outfit, two classic Rasch mean-square fit statistics (Bond et al., 2020), along with RMSD, a more recent measure of item fit (Köhler et al., 2020). A value of 1 represent ideal Infit and Outfit values. Because Infit and Outfit are affected by sample size (Wu & Adams, 2013), we used somewhat tighter guidelines of 1.15 and .85 to flag items for underfit and overfit, respectively. Table 1 and Figure 2 show the Infit and Outfit values. The Infit values were all within the threshold range, but several items showed notable Outfit overfit (e.g., scores were too predictable; items 2, 4, 12, and 15) and several others showed relatively high Outfit values (items 6, 11, 20, 25, 31, 34), which reflect excessively noisy responses that are more problematic for measurement. As we will see later, some of these items were among the “easiest,” most endorsed items in the BICB and showed notable gender-based DIF.

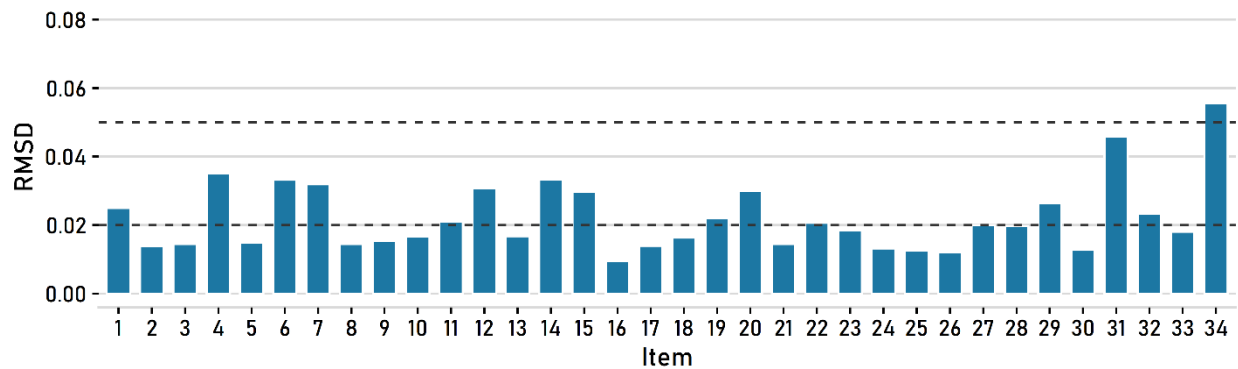
**Figure 2.** Infit and Outfit item fit values for the BICB items.



For the RMSD item fit statistic, Köhler et al. (2020) suggested benchmark values for misfit: negligible ( $\text{RMSD} < .02$ ), small ( $.02 \leq \text{RMSD} < .05$ ), medium ( $.05 \leq \text{RMSD} < .08$ ), and large ( $\text{RMSD} \geq .08$ ). Figure 3 shows the RMSD values with a .05 threshold. Many of the BICB items fell within the “small misfit” range, one item (item 31) neared the medium threshold, and one item (item 34) showed medium misfit. The two items with the largest RMSD values were among the most underfitting items based on Outfit, indicating some consistency between these

fit statistics.

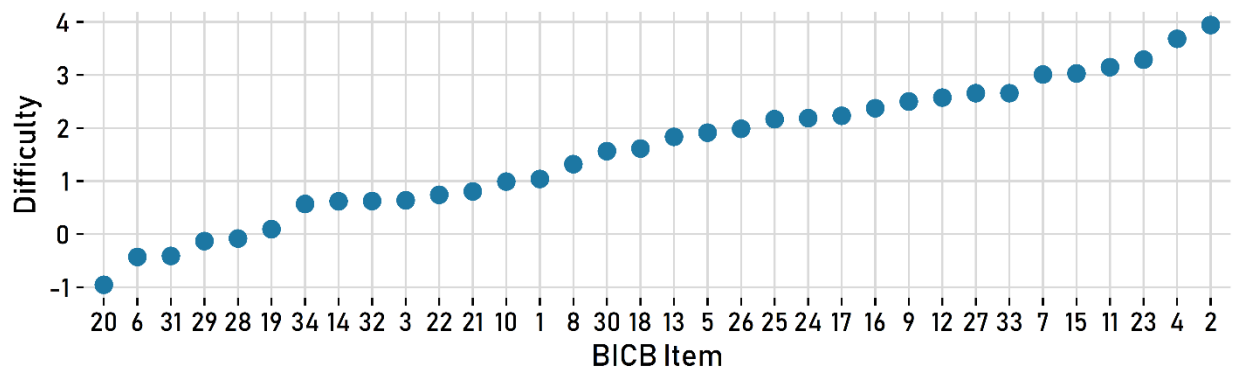
**Figure 3.** RMSD item-fit values for the BICB items.



### 3.4. Item Difficulty Values and Test Information

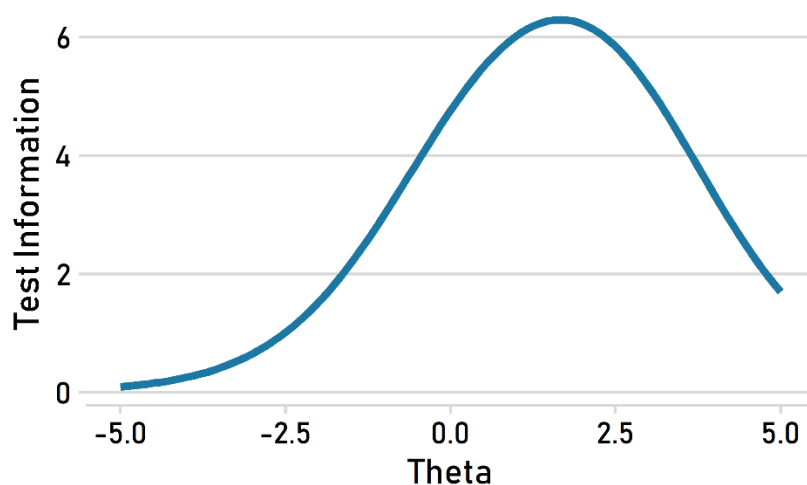
The Rasch model's estimates of the BICB difficulty values suggest that the test is reasonably "hard" and is targeted toward samples with medium and high levels of everyday creativity. As Figure 4 shows, the vast bulk of the items had difficulty values greater than 0. The values ranged from  $-0.96$  (item 20: made someone a present) to  $3.94$  (item 2: wrote a novel). Because the model centers the underlying trait theta scores at 0, the difficulty estimates indicate that, for most of the items, only people with above average levels of everyday creativity are likely to endorse them.

**Figure 4.** Difficulty ( $b$ ) values for the BICB items, sorted easiest to hardest.



The test information function, shown in Figure 5, illustrates the region of the underlying trait that is measured most precisely. For the BICB, the test information peaked at a trait level of around +1.65, consistent with the moderate difficulty values of the scale's items.

**Figure 5.** Test information function for the BICB.



The BICB has a reasonable test information profile for its intended use and population. The scale provides the most information around the middle to the high end of the trait, so it can most reliably sort respondents in that range. Whereas measures of normal personality traits and individual differences usually aim to center their reliability around the middle of the trait region (e.g., Silvia & Rodriguez, 2020), it seems sensible for a measure of everyday creativity to have greater reliability for the higher rather than the lower region of the trait, inasmuch as there is greater interest in understanding and differentiating people higher in creativity than people lower in it.

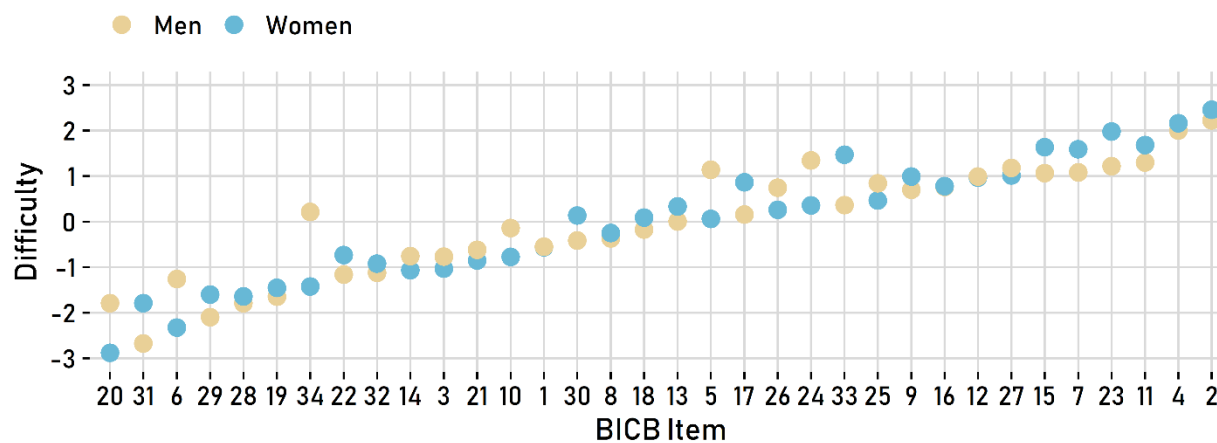
### 3.5. Differential Item Functioning

In Rasch and item response theory models, the probability of an item response should be a function of only people's underlying trait level (Osterlind & Everson, 2009). When members of

different groups have the same trait level but different response probabilities, then the item is said to show differential item functioning (DIF). In the case of two groups, for example, an item with DIF favors one group. Understanding whether a scale's items display DIF is important to establishing that the scale's overall score has the same metric and meaning across groups (Penfield & Camilli, 2006). For the BICB, DIF has not yet been evaluated. We thus explored DIF using Rasch trees (Strobl et al., 2015), a method that uses model-based recursive partitioning to identify DIF. A virtue of this approach is that it can explore DIF for continuous variables, such as age, and identify optimal cut-points from the data. To promote parsimonious Rasch trees, we used a Bonferroni-corrected alpha of .01 and required the nodes (the final groupings of participants based on age) to have at least 400 people.

We first evaluated DIF for gender. The Rasch tree identified significantly different profiles for men and women, which are shown in Figure 6. This figure depicts the estimated difficulty for each item for men and women. (Note that the Rasch model function in *psychotree* uses different identification constraints than the TAM models, so the items' *b* scaling is centered on zero).

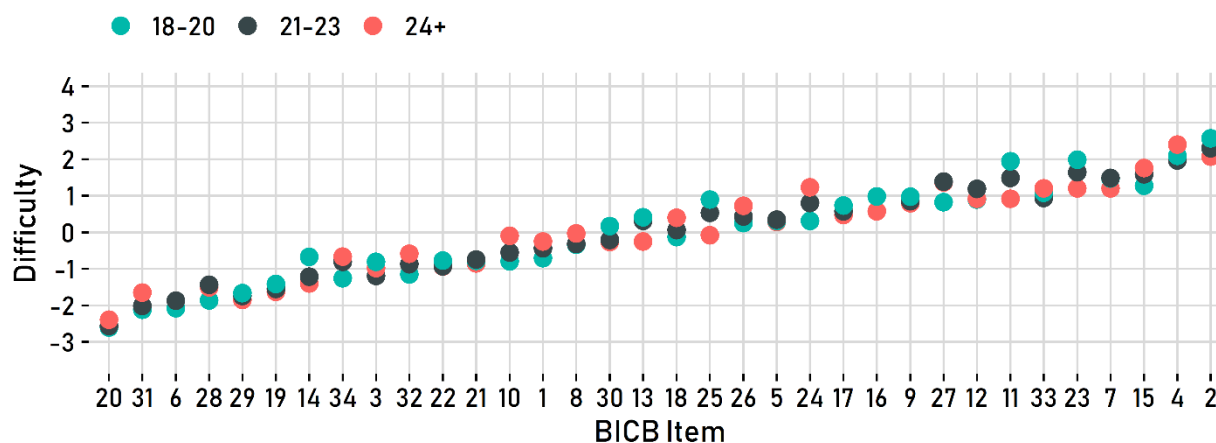
**Figure 6.** Rasch-tree estimated gender profiles in difficulty values for the BICB items.



The figure illustrates that men and women are broadly similar on most BICB items, but there is clear item bias for some of them. Unlike achievement tests, where DIF can indicate unwanted or subtle biases, DIF in an activity scale like the BICB often reflects different cultural norms and affordances that apply to the groups. Many of the BICB items, for example, are culturally gendered, and many of these showed DIF. For example, for items 20 (“Made someone a present”) and 34 (“Made a collage”), women have a much lower difficulty value than men, indicating that, given men and women with equal levels of everyday creativity, it is “easier” for women to endorse that they have made someone a present or made a collage. Some items, however, show DIF but have no obvious gendered quality, such as item 33 (“Composed a piece of music”), for which it was easier for men to endorse. It is worth pointing out that the two items with the worst RMSD fit values (items 31 and 34) and most of the items with the highest outfit values (items 6, 11, 20, 25, 31, 34) showed notable gender-based DIF.

For age DIF, the Rasch tree first branched into two groups: people  $\leq 20$  years old and people  $> 20$  years old. This older group, in turn, was further partitioned, yielding three final nodes: (1) 18-20 years old ( $n = 1202$ ), (2) 21-23 years old ( $n = 692$ ); and (3) 24 and older ( $n = 465$ ). Figure 7 illustrates the findings.

**Figure 7.** Rasch-tree estimated age profiles in difficulty values for the BICB items.



For the most part, age-based DIF was much less striking than gender-based DIF—the three age groups were largely the same. In a handful of cases, the oldest age group (24+ years, shown in red) diverged from the rest. For example, given identical trait scores, people in the older group were nevertheless more likely to endorse item 25 (“Designed and planted a garden”), a kind of creative activity that is less feasible for young college students who are often living in on-campus housing. Likewise, item 11 (“Had an article published”) was more likely to be endorsed by older participants despite holding trait levels constant. Overall, however, the patterns of age DIF seem modest and comprehensible in light of the different interests and affordances for younger and older participants.

#### **4. Discussion**

Our psychometric evaluation of the BICB suggests several strengths as well as some opportunities for future refinement. First, the BICB showed solid dimensionality, viewed as essential versus strict unidimensionality. Factor analysis suggested one dominant, common factor along with at least one minor factor. The secondary, specific factors reflected local dependence—overlap in meaning between relatively redundant item pairs—rather than substantive facets for different domains of creativity, so seeing the BICB as unidimensional is credible. Our dimensionality findings are consistent with past work using confirmatory factor analysis (Silvia et al., 2012) as well as latent class analysis, which suggested that BICB scores sort into levels (classes varying in intensity) instead of nominal classes composed of distinct domains (von Stumm et al., 2011). The scale appears to be better represented in terms of a single dominant factor instead of a group of subfactors or latent classes (see Silvia et al., 2009). If researchers are seeking a measure of everyday creativity that yields a global score, the BICB is a good option.

At the item level, most of the items fit the Rasch model well, with a handful of poorly fitting items. The items showed a broad spread of difficulty, so the BICB offers measurement

information across a wide range of the underlying trait. Aside from a handful of easy items, however, most items required an above-average trait level for likely endorsement, and the BICB's test information function shows that it provides the most information—yields the highest score reliability—in the moderately high region of the trait. This seems like a practical focus for the scale, in that it provides relatively more reliable information for sorting people with relatively higher levels of creativity.

Finally, we evaluated item bias via differential item functioning. The Rasch tree models indicated evidence for gender-based DIF, with some items favoring men and others favoring women. Although gender differences in a variety of general creative ability tests have revealed mixed results at best (Baer & Kaufman, 2008), Kaufman (2006) demonstrated that in a large sample men and women rated themselves higher in areas of creative domains consistent with gender stereotypes. This remained consistent in the present sample, where in most cases, the item bias reflected culturally gendered qualities of the items. Men and women with the same underlying level of everyday creativity are nevertheless exposed to different cultural norms and affordances for creative activities, such as making presents for friends and making collages.

For age-based DIF, relatively modest evidence for DIF was found. We do not wish to make too much of the age findings: our sample had relatively few participants older than 30, and they were perhaps atypical because most of them were enrolled in university psychology courses. Instead, we offer the age analyses as food for thought and as an example of DIF models for continuous variables. Common DIF methods require categorical variables, such as gender or group membership, but many interesting continuous variables could be sources of item bias in creativity assessment (e.g., GPA, socioeconomic status, or personality traits). One virtue of the Rasch tree approach to DIF is that it affords DIF models for continuous variables and empirically identifies optimal cut-points so that researchers needn't draw arbitrary category boundaries (e.g., over or under age 40 or above or below the sample median).

The implications of DIF for scale interpretation and revision can be complex and thus



call for a thoughtful approach (Osterlind & Everson, 2009). For an achievement test (e.g., math knowledge) or personality scales, items flagged for DIF are usually good candidates for omitting and replacing with alternative items from a larger item pool. For activity scales that seek to capture the breadth of activities that people actually do in the real world, however, such as the BICB, the implications of DIF are more nuanced. Removing items with notable DIF would improve item fit and ensure that the total scores for men and women are comparable. At the same time, many high DIF items are popular creative activities with large communities of hobbyists, from making presents to designing gardens, so removing high DIF items sacrifices realistic coverage of the construct's domain for statistical purity. At this point, a reasonable middle ground is for researchers to be circumspect about the meaning of reported differences between men and women and reported correlations with age. To the extent that such effects appear, they will be a mix of real differences in levels of the underlying traits and contaminating influences of item bias.

These findings highlight the complex role of age in measures of creative activities and achievement. Many self-report scales yield cumulative scores—people's activities and achievements to date, usually over their adult lifetime. This imposes correlations between age and creativity scores because older respondents have had more time for achievements to accumulate. Other scales, like the BIBC, use a rolling window—the past 12 months, in this case. These instructions should reduce the influence of age, but many items nevertheless hook into opportunities that come only with age, such as opportunities to plant gardens or publish articles. Choosing to omit or revise these items will need to balance statistical criteria against realistic coverage of the domain of everyday creativity.

Our study has important limitations to consider as well. The generalizability of our sample should be noted for two reasons: a high percentage of our pool is composed of women, and it is largely recruited from college students in the United States. Thus, the everyday creative behaviors of our respondents represent the sorts of activities that appeal to this relatively

distinctive subcultural group. At the same, the large sample size, broad geographic diversity, and long duration of data collection (roughly a decade) contribute to the diversity of our participant pool. Nevertheless, future cross-cultural psychometric work on the BICB would help ensure that suggested scale revisions make it more appropriate for broader use.

Given the popularity of the BICB, it's worth looking ahead to what researchers could do to further refine and improve the scale. Because the scale has found an audience of users in creativity research, it merits a light remodeling to refine its features and update its wording and items. Such a project would require developing and evaluating some new items, of course, but in the meantime, the present analyses suggest that the current BICB could be streamlined. The lowest-hanging fruit are some of the locally dependent items, which impair unidimensionality and add relatively little measurement information. Specifically, we think researchers looking for a slimmer BICB could omit items 8, 17, 19, 23, which are narrower versions of more general items. Omitting some weaker items would create room for adding new ones—perhaps activities related to digital creativity or other activities that were uncommon or didn't exist back when the BICB was first developed. Until it gets remodeled to prepare it for another fruitful decade of research, the BICB appears to be a psychometrically sturdy option for researchers interested in measuring engagement in everyday creative activities.

## 5. References

- Baer, J., & Kaufman, J. C. (2008). Gender differences in creativity. *Journal of Creative Behavior, 42*, 75-105.
- Batey, M. (2007). *A psychometric investigation of everyday creativity*. Unpublished doctoral dissertation. University College, London.
- Batey, M., & Furnham, A. (2008). The relationship between measures of creativity and schizotypy. *Personality and Individual Differences, 45*, 816-821.
- Batey, M., Furnham, A., & Safiullina, X. (2010). Intelligence, general knowledge, and personality as predictors of creativity. *Learning and Individual Differences, 20*, 532-535.
- Benedek, M., Bruckdorfer, R., & Jauk, E. (2020). Motives for creativity: Exploring the what and why of everyday creativity. *Journal of Creative Behavior, 54*, 610-625.
- Bond, T. G., Yan, Z., & Heine, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4<sup>th</sup> ed.). Routledge.
- 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.
- Chen, W. H., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265-289.
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's  $Q_3$ : Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*, 178-194.
- Conner, T. S., DeYoung, C. G., & Silvia, P. J. (2018). Everyday creative activity as a path to flourishing. *Journal of Positive Psychology, 13*, 181-189.
- Cotter, K. N., Christensen, A. P., & Silvia, P. J. (2019). Creativity's role in everyday life. In J. C. Kaufman & R. J. Sternberg (Eds.), *Cambridge handbook of creativity* (2<sup>nd</sup> ed., pp. 640-652). Cambridge University Press.
- Cotter, K. N., & Silvia, P. J. (2019). Ecological assessment in research on aesthetics, creativity,

- and the arts: Basic concepts, common questions, and gentle warnings. *Psychology of Aesthetics, Creativity, & the Arts*, *13*, 211-217.
- Diedrich, J., Jauk, E., Silvia, P. J., Gredlein, J. M., Neubauer, A. C., & Benedek, M. (2018). Assessment of real-life creativity: The Inventory of Creative Activities and Achievements (ICAA). *Psychology of Aesthetics, Creativity, and the Arts*, *12*(3), 304–316.
- Dollinger, S. J. (2003). Need for uniqueness, need for cognition, and creativity. *Journal of Creative Behavior*, *37*, 99–116.
- Elisondo, R. C. (2020). Creative Actions Scale: A Spanish scale of creativity in different domains. *Journal of Creative Behavior*.
- Frith, E., & Loprinzi, P. D. (2020). Is engagement in physical activity and creative activity mutually exclusive? An evaluation of general and domain-specific relationships. *American Journal of Health Studies*, *35*(1), 46-56.
- Furnham, A., Batey, M., Anand, K., & Manfield, J. (2008). Personality, hypomania, intelligence and creativity. *Personality and Individual Differences*, *44*, 1060–1069.
- Fürst, G., & Grin, F. (2018). A comprehensive method for the measurement of everyday creativity. *Thinking Skills and Creativity*, *28*, 84-97.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*, 191-205.
- Hocevar, D. (1979, April). *The development of the Creative Behavior Inventory (CBI)*. Paper presented at the annual meeting of the Rocky Mountain Psychological Association (ERIC Document Reproduction Service No. ED 170 350).
- Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, *76*, 537-549.
- Karwowski, M., Lebeda, I., & Beghetto, R. A. (2019). Creative self-beliefs. In J. C. Kaufman & R. J. Sternberg (Eds.). *Cambridge handbook of creativity* (pp. 396-417). Cambridge

University Press.

- Karwowski, M., Lebuda, I., Szumski, G., & Firkowska-Mankiewicz, A. (2017). From moment-to-moment to day-to-day: Experience sampling and diary investigations in adults' everyday creativity. *Psychology of Aesthetics, Creativity, and the Arts, 11*, 309-324.
- Karwowski, M., Lebuda, I., & Wisniewska, E. (2018). Measuring creative self-efficacy and creative personal identity. *International Journal of Creativity and Problem Solving, 28*, 45-57.
- Kaufman, J. C. (2006). Self-reported differences in creativity by ethnicity and gender. *Applied Cognitive Psychology, 20*, 1065-1082.
- Kaufman, J. C. (2019). Self-assessments of creativity: Not ideal, but better than you think. *Psychology of Aesthetics, Creativity, and the Arts, 13*, 187-192.
- Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The Four C model of creativity. *Review of General Psychology, 13*, 1-12.
- Köhler, C., Robitzsch, A., & Hartig, J. (2020). A bias-corrected RMSD item fit statistic: An evaluation and comparison to alternatives. *Journal of Educational and Behavioral Statistics, 45*, 251-273.
- Marais, I. (2013). Local dependence. In K. B. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 111-130). John Wiley & Sons.
- McKay, A. S., Karwowski, M., & Kaufman, J. C. (2017). Measuring the muses: Validating the Kaufman Domains of Creativity Scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts, 11*, 216-230.
- Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning*. Sage.
- Penfield, R. D., & Camilli, G. (2006). Differential item functioning and item bias. *Handbook of Statistics, 26*, 125-167.
- Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relationships of personality to differential conceptions of creativity: A systematic review. *Psychology of Aesthetics, Creativity, and*

*the Arts*, 11, 59–68.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Downloaded from <https://www.R-project.org>

Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research*. R package version 2.0.9. <https://CRAN.R-project.org/package=psych>

Richards, R. (2007). Everyday creativity: Our hidden potential. In R. Richards (Ed.), *Everyday creativity and new views of human nature: Psychological, social, and spiritual perspectives* (pp. 25–53). American Psychological Association.

Richards, R. (2010). Everyday creativity: Process and way of life—four key issues. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 189-215). Cambridge University Press.

Robitzsch, A., Kiefer, T., & Wu, M. (2020). *TAM: Test Analysis Modules*. R package version 3.6.8. Downloaded from <https://CRAN.R-project.org/package=TAM>

Said-Metwaly, S., Van den Noortgate, W., & Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. *Creativity: Theories-Research-Applications*, 4(2), 238-275.

Silvia, P. J., & Cotter, K. N. (2021). *Researching daily life: A guide to experience sampling and daily diary methods*. American Psychological Association.

Silvia, P. J., Cotter, K. N., & Christensen, A. P. (2017). The creative self in context: Experience sampling and the everyday ecology of creativity. In M. Karwowski & J. C. Kaufman (Eds.), *The creative self: Effect of beliefs, self-efficacy, mindset, and identity* (pp. 275-288). Elsevier.

Silvia, P. J., Eddington, K. M., Harper, K. L., Burgin, C. J., & Kwapil, T. R. (2020). Depressive anhedonia and creative self-concepts, behaviors, and achievements. *Journal of Creative Behavior*.

Silvia, P. J., Kaufman, J. C., & Pretz, J. E. (2009). Is creativity domain-specific? Latent class

- models of creative accomplishments and creative self-descriptions. *Psychology of Aesthetics, Creativity, and the Arts*, 3, 139-148.
- 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.
- Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research*, 102, 443-461.
- Strobl, C., Kopf, J., & Zeileis, A. (2015). Rasch trees: A new method for detecting differential item functioning in the Rasch model. *Psychometrika*, 80(2), 289-316.
- Tempest, G. D., & Radel, R. (2019). Put on your (fNIRS) thinking cap: Frontopolar activation during augmented state creativity. *Behavioural Brain Research*, 373, 112082.
- Velicer, W. (1976) Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
- von Stumm, S., Chung, A., & Furnham, A. (2011). Creative ability, creative ideation, and latent classes of creative achievement: What is the role of personality? *Psychology of Aesthetics, Creativity, and the Arts*, 5(2), 107-114.
- West, M. J., & Somer, E. (2020). Empathy, emotion regulation, and creativity in immersive and maladaptive daydreaming. *Imagination, Cognition and Personality*, 39(4), 358-373.
- Wigert, B., Reiter-Palmon, R., Kaufman, J. C., Silvia, P. J. (2012). Perfectionism: The good, the bad, and the creative. *Journal of Research in Personality*, 46, 775-779.
- Wu, M., & Adams, R. J. (2013). Properties of Rasch residual fit statistics. *Journal of Applied Measurement*, 14, 339-355.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8, 125-145.
- Zeileis, A., Strobl, C., Wickelmaier, F., Komboz, B., & Kopf, J. (2020). *psychotree: Recursive*

*partitioning based on psychometric models*. R package version 0.15.3. Downloaded  
from <https://CRAN.R-project.org/package=psychotree>



Table 1

*Biographical Inventory of Creative Behaviors (BICB): Item Statistics*

Item	Difficulty ( <i>b</i> )	Outfit	Infit	RMSD	Local Dependence
1. Wrote short story	1.04	1.05	1.03	.025	
2. Wrote novel	3.94	.76	.95	.014	4
3. Organized event	.64	.96	.98	.014	
4. Produced script	3.68	.75	.90	.035	2, 27
5. Designed textile	1.91	.91	.98	.015	
6. Decorated room	-.43	1.11*	1.02	.033	
7. Invented product	3.01	.86	.92	.032	
8. Drew cartoon	1.32	.97	.99	.014	10
9. Started club	2.50	.93	.97	.015	
10. Made picture	.99	.99	1.01	.017	8
11. Published article	3.15	1.14	1.01	.021	23
12. Made sculpture	2.57	.78*	.92	.031	
13. Criticized scientific theory	1.84	1.04	1.03	.017	17
14. Made recipes	.62	1.05	1.06*	.033	
15. Produced short film	3.03	.71*	.92	.030	
16. Made webpage	2.37	.95	.99	.009	
17. Created a theory	2.23	.93	.99	.014	13
18. Invented game	1.61	.89	.96	.016	
19. Chosen to lead	.09	.96	.97	.022	32
20. Made a present	-.96	1.15*	.98	.030	
21. Wrote poem	.81	1.05	1.03	.015	
22. Adapted object	.74	.92*	.97	.020	
23. Published research	3.29	1.05	.97	.018	11
24. Choreographed dance	2.19	.95	1.01	.013	
25. Designed garden	2.17	1.11	1.00	.013	
26. Made photography portfolio	1.98	.90	.98	.012	
27. Acted	2.66	.88	.95	.020	4
28. Gave speech	-.08	1.07*	1.03	.020	
29. Mentored others	-.13	.99	.97	.026	
30. Designed experiment	1.57	.92	.97	.013	
31. Wrote jokes	-.41	1.20*	1.09*	.046	
32. Served as leader	.62	1.02	1.03	.023	19
33. Composed music	2.66	1.03	.98	.018	
34. Made collage	.57	1.18*	1.12*	.056	

**Note.** The item labels are abbreviated stems, not the actual items. The scale is available for download at OSF (<https://osf.io/4s9p6/>). Infit and Outfit values with an asterisk have significant *t*-values greater than |2.0|. Locally dependent items have a  $Q_3$  correlations greater than |.20|.