



Measuring Creative Self-Efficacy: An Item Response Theory Analysis of the Creative Self-Efficacy Scale

Amy Shaw^{1*}, Melissa Kapnek² and Neil A. Morelli³

¹Department of Psychology, Faculty of Social Sciences, University of Macau, Avenida da Universidade, Macau, China,

²Berke Assessment, Atlanta, GA, United States, ³Codility, San Francisco, CA, United States

OPEN ACCESS

Edited by:

Laura Galiana,
University of Valencia, Spain

Reviewed by:

Irene Fernández,
University of Valencia, Spain
Matthias Trendel,
Technical University of Dortmund,
Germany

*Correspondence:

Amy Shaw
amyshaw@um.edu.mo

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 08 March 2021

Accepted: 04 June 2021

Published: 01 July 2021

Citation:

Shaw A, Kapnek M and
Morelli NA (2021) Measuring Creative
Self-Efficacy: An Item Response
Theory Analysis of the Creative
Self-Efficacy Scale.
Front. Psychol. 12:678033.
doi: 10.3389/fpsyg.2021.678033

Applying the graded response model within the item response theory framework, the present study analyzes the psychometric properties of Karwowski's creative self-efficacy (CSE) scale. With an ethnically diverse sample of US college students, the results suggested that the six items of the CSE scale were well fitted to a latent unidimensional structure. The scale also had adequate measurement precision or reliability, high levels of item discrimination, and an appropriate range of item difficulty. Gender-based differential item functioning analyses confirmed that there were no differences in the measurement results of the scale concerning gender. Additionally, openness to experience was found to be positively related to the CSE scale scores, providing some support for the scale's convergent validity. Collectively, these results confirmed the psychometric soundness of the CSE scale for measuring CSE and also identified avenues for future research.

Keywords: creativity, creative self-efficacy, personality, creative self-efficacy scale, item response theory

INTRODUCTION

Defined as “the belief one has the ability to produce creative outcomes” (Tierney and Farmer, 2002, p. 1138), creative self-efficacy (CSE; Tierney and Farmer, 2002, 2011; Beghetto, 2006; Karwowski and Barbot, 2016) has attracted increasing attention in the field of creativity research. The concept of CSE originates from and represents an elaboration of Bandura's (1997) self-efficacy construct. According to Bandura (1997), self-efficacy influences what a person tries to accomplish and how much effort she/he may exert on the process. As such, CSE reflects a self-judgment of one's own creative capabilities or potential which, in turn, affects the person's activity choice and effort and, ultimately, the attainment of innovative outcomes. Lemons (2010) even claimed that it is not the competence itself but the mere belief about it that matters. Therefore, CSE appears to be an essential psychological attribute for researchers to understand the exhibition and improvement of creative performance. Indeed, there has been empirical evidence supporting the motivational importance of CSE and its capability of predicting crucial performance outcomes in both educational and workplace contexts (e.g., Schack, 1989; Tierney and Farmer, 2002, 2011; Choi, 2004; Beghetto, 2006; Gong et al., 2009; Karwowski, 2012, 2014; Karwowski et al., 2013; Puente-Díaz and Cavazos-Arroyo, 2017).

Given the important role of CSE, having psychometrically sound assessments of this construct is critical. Responding to the call for more elaborate CSE measures (Beghetto, 2006; Karwowski, 2011),

the Short Scale of Creative Self (SSCS; Karwowski, 2012, 2014; Karwowski et al., 2018) was designed to measure trait-like CSE and creative personal identity (CPI; the belief that creativity is an important element of self-description; Farmer et al., 2003) by asking respondents to indicate the degree to which they include the construct as part of who they are on a 5-point Likert scale. The SSCS is composed of 11 items with six items measuring CSE and five items measuring CPI; CSE is often studied together with CPI, but both of the CSE and CPI subscales can be used as stand-alone scales (Karwowski, 2012, 2014; Karwowski et al., 2018). Specifically, CSE is described by the following six statements on the SSCS: Item (3) “I know I can efficiently solve even complicated problems,” Item (4) “I trust my creative abilities,” Item (5) “Compared with my friends, I am distinguished by my imagination and ingenuity,” Item (6) “I have proved many times that I can cope with difficult situations,” Item (8) “I am sure I can deal with problems requiring creative thinking,” and Item (9) “I am good at proposing original solutions to problems” (Karwowski et al., 2018, p. 48).

Since its introduction, the CSE scale has attracted research attention and there is some validity evidence supporting its use. In the formal scale development and validation study based on a sample of $n = 622$ participants, Karwowski et al. (2018) found that the 6-item CSE scale had a very good internal consistency reliability estimate, consisted of one predominant factor (i.e., CSE) and showed good convergent validity with moderate to large correlations with other CSE measures, such as the brief measures proposed by Beghetto (2006) and Tierney and Farmer (2002). Other empirical studies that adopted this instrument also suggested that the scale possessed fairly good estimates of reliability and validity (e.g., Karwowski, 2012, 2014, 2016; Karwowski et al., 2013; Liu et al., 2017; Puente-Díaz and Cavazos-Arroyo, 2017; Royston and Reiter-Palmon, 2019; Qiang et al., 2020).

Despite its promise, few studies beyond those by the scale developers have been conducted to investigate psychometric soundness of the 6-item CSE scale in terms of reliability and validity. Moreover, although the CSE scale has been thoroughly examined in samples from Poland (e.g., Karwowski et al., 2018), it has not yet been investigated for its psychometric properties in the US sample. Finally, all psychometric studies of the CSE scale so far have relied on the classical test theory (CTT) approaches in lieu of more appropriate modern test theory or item response theory (IRT; Steinberg and Thissen, 1995; Embretson and Reise, 2000) approaches (see also Shaw, Elizondo, and Wadlington 2020; for a recent discussion of applying advanced IRT models).

This study thus attempts to remedy these issues. Within the validity framework established by the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014), here we report on a psychometric evaluation of the CSE scale in a sample of the US college students using IRT analyses. Specifically, item quality, measurement precision or reliability, dimensionality, and relations to external variables are evaluated. To our knowledge, it is also the first study that applies IRT to investigate the latent trait and item-level characteristics, such as item difficulty and discrimination of

the CSE scale. Differential item functioning (DIF) analysis was conducted to examine the equivalence of individual item functioning across two gender subgroups, given some empirical findings on gender differences (albeit weak and inconsistent) in CSE (higher self-rated CSE by males; Beghetto, 2006; Furnham and Bachtiar, 2008; Karwowski, 2009). Additionally, concurrent validity was examined *via* evaluating the relationships among the CSE scale scores and the Big Five personality traits that have been found to be linked to CSE positively or negatively (e.g., Karwowski et al., 2013).

MATERIALS AND METHODS

Participants

A total of $n = 173$ undergraduates at a large public university in the southern United States participated in this study for research credits. Participants' ages ranged from 18 to 24 years with an average age of 20.60 ($SD = 0.80$). Among these subjects, $n = 101$ (58.4%) were female and $n = 72$ (41.6%) were male. The most commonly represented majors in the sample were psychology (32.8%), other social sciences (28.1%), and engineering (23.2%). Based on self-declared demographic information, 38.5% were Hispanic/Latino, 26.3% were Caucasian, 20.7% were Asian, 11.2% were African-American, and 3.4% selected other for ethnicity; the sample was thus ethnically diverse.

Study Procedure and Materials

After providing their written informed consent, participants completed a standard demographic survey in addition to the 6-item CSE scale (Karwowski, 2012, 2014; Karwowski et al., 2018) and the Ten-Item Personality Inventory (TIPI; Gosling et al., 2003) for Big Five personality. The TIPI is comprised of 10 items with each containing a pair of trait descriptors; each trait is represented by two items: one stated in a way that characterizes the positive end of the trait and the other stated in a way that characterizes the negative end. For the TIPI, participants were asked to rate the extent to which the pair of traits applies to him/her on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). All the TIPI trait scales showed similar internal consistency estimates to those reported in other studies (Gosling et al., 2003; Muck et al., 2007; Romero et al., 2012; Laguna et al., 2014; Azkhosh et al., 2019): Extraversion ($\alpha = 0.68, \omega = 0.69$), Agreeableness ($\alpha = 0.45, \omega = 0.47$), Conscientiousness ($\alpha = 0.51, \omega = 0.52$), Emotional Stability ($\alpha = 0.71, \omega = 0.71$), and Openness to Experience ($\alpha = 0.46, \omega = 0.47$). These values are relatively low according to the rule of thumb of $\alpha = 0.70$ (Nunnally, 1978, p. 245) but considered reasonably acceptable for a scale of such brevity (Gosling et al., 2003; Romero et al., 2012). For the CSE scale, participants were asked to indicate the extent to which each of the statements describes him/her on a 5-point Likert scale (1 = definitely not; 5 = definitely yes). Therefore, possible total scores on the CSE scale could range from 6 to 30, with higher scores indicating greater CSE.

TABLE 1 | Slope and category threshold parameter estimates for all six Items.

| Item | Label | Slope (a) | s.e. | Threshold 1 (b_1) | s.e. | Threshold 2 (b_2) | s.e. | Threshold 3 (b_3) | s.e. | Threshold 4 (b_4) | s.e. |
|------|--------|---------------|------|-----------------------|------|-----------------------|------|-----------------------|------|-----------------------|------|
| 1 | SSCS 3 | 1.84 | 0.31 | -1.15 | 0.18 | -0.09 | 0.11 | 0.71 | 0.14 | 1.71 | 0.24 |
| 2 | SSCS 4 | 1.18 | 0.21 | -2.41 | 0.41 | -0.87 | 0.20 | 0.30 | 0.15 | 2.23 | 0.38 |
| 3 | SSCS 5 | 1.40 | 0.24 | -2.45 | 0.37 | -1.39 | 0.23 | -0.22 | 0.13 | 1.55 | 0.25 |
| 4 | SSCS 6 | 1.15 | 0.22 | -3.42 | 0.62 | -1.34 | 0.27 | -0.16 | 0.16 | 1.72 | 0.31 |
| 5 | SSCS 8 | 1.28 | 0.24 | -1.41 | 0.26 | -0.30 | 0.15 | 1.07 | 0.21 | 2.33 | 0.39 |
| 6 | SSCS 9 | 1.65 | 0.28 | -1.81 | 0.27 | -0.59 | 0.15 | 0.40 | 0.13 | 1.48 | 0.22 |

Note. SSCS = Short Scale of Creative Self.

TABLE 2 | Marginal fit (χ^2) and standardized LD χ^2 statistics.

| Item | Label | Marginal χ^2 | Standardized LD χ^2 of item pairs | | | | | |
|------|--------|-------------------|--|--------|--------|--------|--------|--|
| | | | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | |
| 1 | SSCS 3 | 0.2 | - | | | | | |
| 2 | SSCS 4 | 0.2 | 1.3 | - | | | | |
| 3 | SSCS 5 | 0.2 | 2.3 | 2.5 | - | | | |
| 4 | SSCS 6 | 0.2 | 0.2 | 2.7 | -0.4 | - | | |
| 5 | SSCS 8 | 0.2 | 0.1 | 2.9 | 1.0 | 1.8 | - | |
| 6 | SSCS 9 | 0.1 | 1.6 | 1.0 | 0.0 | 2.1 | -0.2 | |

Note. SSCS = Short Scale of Creative Self.

ANALYSES AND RESULTS

All cases were included in the final analyses ($n = 173$). The unidimensional IRT analysis of the CSE scale was conducted in IRTPRO (Cai et al., 2011) using Samejima's (1969, 1997) graded response model (GRM), a suitable IRT model for data with ordered polytomous response categories, such as Likert-scale survey data (Steinberg and Thissen, 1995; Gray-Little et al., 1997). In GRM, each item has a slope parameter and between-category threshold parameters (one less than the number of response categories). In the current analysis, each item had five ordered response categories and thus four threshold parameters. Typically, in IRT models, the latent trait scale (theta-axis) is set with the assumption that the sample group is from a normally distributed population (mean value = 0; standard deviation = 1). This also applies to the GRM in the current study, and therefore, a theta value of 0 represents average CSE and a theta value of -1.00, for example, suggests being one standard deviation below the average.

Table 1 lists the item parameter estimates for all six items together with their standard errors, which can be used to evaluate how each item performs. The slope or item discrimination parameter $a(s)$ reflects the strength of the relationship between the item response and the underlying construct, which indicates how fast the probabilities of responses change across the trait level (i.e., CSE). Generally, items with higher slope parameters provide more item information. The slopes for the six items were all higher than 1.00, and the associated standard errors ranged from 0.21 to 0.31, indicating a satisfactory degree of discriminating power for the six items (Steinberg and Thissen, 1995; Cai et al., 2011). The category threshold or item boundary location parameters $b(s)$ reflect the points on the latent trait

scale (theta-axis) at which a respondent has a 50% probability of endorsing above the threshold; higher threshold parameters suggest the items are more difficult (i.e., requiring higher trait level to endorse). For example, looking at the first row of **Table 1**, one can see that for Item 1 (or Item 3 on the original SSCS scale: "I know I can efficiently solve even complicated problems"), a respondent with a trait level (theta value) of -1.15 (b_1) has a 50% probability of endorsing "2 = probably not" or higher, with a trait level of -0.09 (b_2) has a 50% probability of endorsing "3 = possibly" or higher, with a trait level of 0.71 (b_3) has a 50% probability of endorsing "4 = probably yes" or higher, and with a trait level of 1.71 (b_4) has a 50% probability of endorsing "5 = definitely yes." As displayed in **Table 1**, all threshold parameter estimates ranged from -3.42 to 2.33, indicating that the items provided good measurement in terms of item difficulty across an adequate range of the underlying trait (i.e., CSE).

One assumption underlying the application of unidimensional IRT models is that a single psychological continuum (i.e., the latent trait) accounts for the covariation among the responses. The assumption of unidimensionality and the model fit could be evaluated simultaneously by examining the presence of local dependence (LD) among pairs of the scale items. Referring to excess covariation between item pairs that could not be accounted for by the single latent trait in the unidimensional IRT model, LD implies that the model is not adequately capturing all item covariances. The standardized chi-square statistic (standardized LD χ^2 ; Chen and Thissen, 1997) was used for the evaluation of LD; standardized LD χ^2 values of 10 or greater are generally considered noteworthy. Goodness of fit of the GRM was evaluated using the M_2 statistic and the associated RMSEA value (Cai et al., 2006; Maydeu-Olivares and Joe, 2006). As presented in **Table 2**, the largest standardized LD χ^2 value was 2.9 (less than 10)

so there was no indication of LD among the six items and thus no violation of unidimensionality for pairs of items. Goodness of fit indices also demonstrated that the unidimensional GRM had satisfactory fit [$M_2(305) = 438.22, p < 0.001; RMSEA = 0.04$]. Besides, we looked at the summed-score-based item fit statistics [$S-\chi^2$ item-level diagnostic statistics; Orlando and Thissen, 2000, 2003; also see Roberts (2008) for a discussion of extensions] for further evaluation (significant values of p indicate lack of item fit). As presented in **Table 3**, all probabilities were above 0.05 so that there was no item flagged as potentially problematic or misfitting.

In **Figure 1**, the left and right panels present the test information curve (together with its corresponding standard errors line) and test characteristic curve, respectively. The test information curve was created by adding together all six-item information curves. The test information curve describes varying measurement precision provided at each trait level (IRT information is the expected value of the inverse of the error variances for each estimated value of the latent trait) and estimates how well the

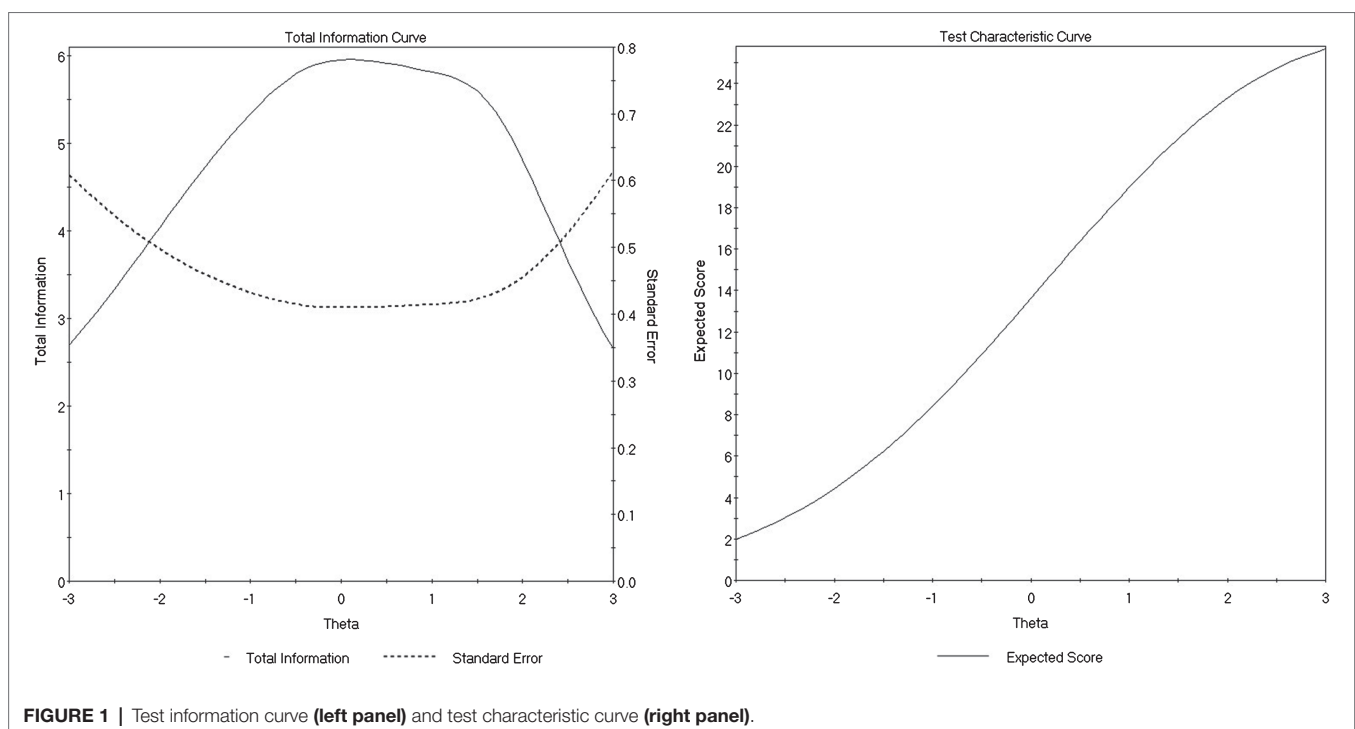
construct is measured at all levels of the underlying trait, thus showing how well the measure functions as a whole across the latent trait continuum for the model. Generally, more psychometric information equals greater measurement precision (with lower error). As graphically illustrated in the left panel of **Figure 1**, the test information curve peaks in the middle (total information for the entire scale is approximately 6.00 in that range), indicating that the test provided the most information (or smallest standard errors of measurement) in the middle (and slightly-to-the-right) range of trait level estimates (where most of the respondents are located) but little information for those at extremely low or high ends of CSE (i.e., theta values outside the range of -3.00 to 3.00 along the construct continuum). The calculated *Expected A Posteriori*-based marginal reliability value was 0.82. Thus, the CSE scale in the current study appeared to work well (and was the most informative/sensitive) for differentiating individuals in the middle and middle-to-high of the trait range (where most people reside). The test characteristic curve, as displayed in the right panel, presents the expected values of the summed observed scores of the entire scale as a function of theta values (i.e., the CSE trait levels). For instance, the zero-theta value corresponds to the expected summed score of 13.63. The close-to-linear curve for values of CSE on the continuum between -2.00 and 2.00 suggests that the summed observed scores were a good approximation of the latent trait scores estimated in GRM.

Differential item functioning detection was performed using the Mantel test (Mantel, 1963). No evidence of DIF was found between male and female respondents [DIF contrasts were below 0.50, Mantel-Haenszel probabilities for all items were above 0.05, and thus, there was no indication of a statistically

TABLE 3 | $S-\chi^2$ item-level diagnostic statistics.

| Item | Label | χ^2 | d.f. | Probability |
|------|--------|----------|------|-------------|
| 1 | SSCS 3 | 48.13 | 41 | 0.21 |
| 2 | SSCS 4 | 50.17 | 42 | 0.18 |
| 3 | SSCS 5 | 50.90 | 40 | 0.12 |
| 4 | SSCS 6 | 43.59 | 39 | 0.28 |
| 5 | SSCS 8 | 46.37 | 43 | 0.34 |
| 6 | SSCS 9 | 31.37 | 40 | 0.83 |

Note. SSCS, Short Scale of Creative Self.



significant difference of item functioning across gender subgroups; the effect sizes of all items were also classified as small/negligible according to the ETS delta scale (Zieky, 1993; Sireci and Rios, 2013; Shaw et al., 2020)], suggesting item fairness of the CSE scale regarding gender.

In addition, concurrent validity was examined by evaluating the Big Five personality correlates of the CSE total scores using Pearson's correlation. Replicating part of the findings in past work (e.g., Jaussi et al., 2007; Silvia et al., 2009; Karwowski et al., 2013), openness to experience was found to be positively related to CSE ($r = 0.23$, $p < 0.01$). Other traits, however, were not found to be related to CSE in the current sample: Extraversion ($r = 0.13$, $p = 0.09$), Agreeableness ($r = 0.08$, $p = 0.30$), Conscientiousness ($r = 0.10$, $p = 0.19$), and Emotional Stability ($r = 0.06$, $p = 0.43$).

DISCUSSION

In the present study, we aimed to better understand the psychometric properties of the 6-item CSE scale (Karwowski, 2012, 2014; Karwowski et al., 2018). Applying GRM in IRT, we found that the items were well fitted to a single latent construct model, providing support for the scale as a unidimensional measure of CSE. This finding is in line with previous studies using more traditional but less sophisticated approaches for categorical response data in CTT (e.g., Karwowski et al., 2018). The IRT analyses also suggested high levels of item discrimination, an appropriate range of item difficulty, as well as satisfactory measurement precision primarily suitable for respondents near average CSE. Furthermore, the gender-based DIF detection confirmed that there was no gender DIF for the CSE items, so that any score difference between the two gender subgroups on the CSE scale could be attributable to meaningful differences in the underlying construct (i.e., CSE), making the CSE scale, a useful instrument for studying gender and CSE. Regarding correlations with relevant criteria, CSE positively related to openness to experience, exhibiting some convergence validity. Collectively, these results provided initial evidence supporting the psychometric soundness of the CSE scale for measuring CSE among the US college students.

Notably, the CSE scale is a domain-general self-rating scale. Despite the ongoing debate on whether creativity and creative self-concept shall be better measured as domain-general or domain-specific constructs, Pretz and McCollum (2014) suggested that there is an association between CSE and the belief to be creative on both domain-general and more domain-specific self-ratings, albeit the varying effect sizes that might be dependent on domain specificity. Moreover, in spite of some doubts concerning self-rated creativity as a valid and useful measure of actual creativity (Reiter-Palmon et al., 2012), there is research evidence suggesting that subjective and objective ratings of creativity tended to be positively correlated (Furnham et al., 2005); a growing body of empirical work in the CSE literature has also elucidated that self-judgments about one's creative potential could serve as a crucial motivational/volitional factor driving

actions that may lead to creative performance (Tierney and Farmer, 2002, 2011; Farmer et al., 2003; Beghetto, 2006; Karwowski and Barbot, 2016). At the very least, self-assessments of creativity could be a nice complement to other types of creativity assessments in cases where objective performance metrics are unavailable for research.

Several limitations of the current study are also worth noting. First, the relatively small sample size makes all interpretations of the results subject to suspicion, given the fact that a GRM was applied and each item had five response categories—the large amount of possible response patterns definitely benefits from having a larger sample size which would allow for a more convincing conclusion. Second, although an ethnically diverse sample was used, it was a convenience college student sample, and thus, the results should be considered within the context, and any generalization of the findings to other populations shall be done with caution. That said, further research with larger and more representative college student sample or samples from other populations (e.g., working adults, graduate students, and high school students) is warranted. Third, even though no gender difference in the CSE scale scores was observed in the current study, this finding should be interpreted with caution given the fact that the sample was slightly predominated by females. Also with the sample consisting of a majority of students from psychology or other social sciences majors, the results regarding the absence of gender differences in the current sample require further examination. Studies have not converged on the relationship between CSE and gender, but in a study by Kaufman (2006), males self-reported greater creativity than females in areas of science and sports, whereas females self-reported greater creativity than males in domains of social communication and visual artistic factors. Therefore, it is likely that the characteristics of our current sample (predominated by females and mostly in social sciences) limited the capacity of the study to detect potential gender differences in CSE. Future research using more gender-balanced samples with diverse academic majors is recommended. Last, the inherent limitation of the personality scale used in the current study may have contributed to the smaller size of the CSE-openness correlation compared to findings in other studies that used more comprehensive personality measures (e.g., Furnham et al., 2005; Karwowski et al., 2013; Pretz and McCollum, 2014). Although the TIPI has been widely used and is characterized by satisfactory correlations with other personality measures, this brief personality scale only consists of 10 items (two for each trait) which often inevitably results in lower internal consistencies and somewhat diminished validities (Gosling et al., 2003; Romero et al., 2012).

In sum, by demonstrating satisfactory item-level discriminating power, an appropriate range of item difficulty, good item fit and functioning, adequate measurement precision or reliability, and unidimensionality for the CSE scale, this study provided support for its internal construct validity. The positive CSE-openness relationship finding also provided some evidence for the scale's convergent validity. Future research may further

assess the predictive validity of the CSE scale on outcome measures, preferably in comparison with other less elaborate measures of CSE in the literature. Based on the results of the current work, the 6-item CSE scale could be a useful and appropriate CSE measurement tool for researchers and practitioners to conveniently incorporate in studies. It is also our hope that this study together with past work will facilitate even more efforts to develop, validate, and refine instruments for CSE.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are subject to the following licenses/restrictions: Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the authors upon reasonable request. Requests to access these datasets should be directed to AS, amyshaw@um.edu.mo.

REFERENCES

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Azkhosh, M., Sahaf, R., Rostami, M., and Ahmadi, A. (2019). Reliability and validity of the 10-item personality inventory among older Iranians. *Psychol. Russia: State Art* 12, 28–38. doi: 10.11621/pir.2019.0303
- Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York, NY: Freeman.
- Beghetto, R. A. (2006). Creative self-efficacy: correlates in middle and secondary students. *Creat. Res. J.* 18, 447–457. doi: 10.1207/s15326934crj1804_4
- Cai, L., du Toit, S. H. C., and Thissen, D. (2011). *IRTPRO: Flexible, Multidimensional, Multiple Categorical IRT Modeling [Computer software]*. Chicago, IL: Scientific Software International.
- Cai, L., Maydeu-Olivares, A., Coffman, D. L., and Thissen, D. (2006). Limited-information goodness-of-fit testing of item response theory models for sparse 2^p tables. *Br. J. Math. Stat. Psychol.* 59, 173–194. doi: 10.1348/000711005X66419
- Chen, W. H., and Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *J. Educ. Behav. Stat.* 22, 265–289. doi: 10.3102/10769986022003265
- Choi, J. N. (2004). Individual and contextual predictors of creative performance: the mediating role of psychological processes. *Creat. Res. J.* 16, 187–199. doi: 10.1080/10400419.2004.9651452
- Embretson, S. E., and Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Farmer, S. M., Tierney, P., and Kung-McIntyre, K. (2003). Employee creativity in Taiwan: an application of role identity theory. *Acad. Manag. J.* 46, 618–630. doi: 10.5465/30040653
- Furnham, A., and Bachtiar, V. (2008). Personality and intelligence as predictors of creativity. *Personal. Individ. Differ.* 45, 613–617. doi: 10.1016/j.paid.2008.06.023
- Furnham, A., Zhang, J., and Chamorro-Premuzic, T. (2005). The relationship between psychometric and self-estimated intelligence, creativity, personality and academic achievement. *Imagin. Cogn. Pers.* 25, 119–145. doi: 10.2190/530V-3M9U-7UQ8-FMBG
- Gong, Y., Huang, J., and Farh, J. (2009). Employee learning orientation, transformational leadership, and employee creativity: the mediating role of creative self-efficacy. *Acad. Manag. J.* 52, 765–778. doi: 10.5465/amj.2009.43670890
- Gosling, S. D., Rentfrow, P. J., and Swann, W. B. Jr. (2003). A very brief measure of the Big-Five personality domains. *J. Res. Pers.* 37, 504–528. doi: 10.1016/S0092-6566(03)00046-1
- Gray-Little, B., Williams, V. S. L., and Hancock, T. D. (1997). An item response theory analysis of the Rosenberg Self-Esteem Scale. *Personal. Soc. Psychol. Bull.* 23, 443–451. doi: 10.1177/0146167297235001
- Jaussi, K. S., Randel, A. E., and Dionne, S. D. (2007). I am, I think I can, and I do: the role of personal identity, self-efficacy, and cross-application of experiences in creativity at work. *Creat. Res. J.* 19, 247–258. doi: 10.1080/10400410701397339
- Karwowski, M. (2009). I'm creative, but am I creative? Similarities and differences between self-evaluated small and big-C creativity in Poland. *Korean J. Thinking Probl. Solving* 19, 7–29. doi: 10.5964/ejop.v8i4.513
- Karwowski, M. (2011). It doesn't hurt to ask... But sometimes it hurts to believe: Polish students' creative self-efficacy and its predictors. *Psychol. Aesthet. Creat. Arts* 5, 154–164. doi: 10.1037/a0021427
- Karwowski, M. (2012). Did curiosity kill the cat? Relationship between trait curiosity, creative self-efficacy and creative role identity. *Eur. J. Psychol.* 8, 547–558.
- Karwowski, M. (2014). Creative mindsets: measurement, correlates, consequences. *Psychol. Aesthet. Creat. Arts* 8, 62–70. doi: 10.1037/a0034898
- Karwowski, M. (2016). The dynamics of creative self-concept: changes and reciprocal relations between creative self-efficacy and creative personal identity. *Creat. Res. J.* 28, 99–104. doi: 10.1080/10400419.2016.1125254
- Karwowski, M., and Barbot, B. (2016). "Creative self-beliefs: their nature, development, and correlates," in *Cambridge Companion to Reason and Development*. eds. J. C. Kaufman and J. Baer (New York, NY: Cambridge University Press), 302–326.
- Karwowski, M., Lebuda, I., and Wiśniewska, E. (2018). Measuring creative self-efficacy and creative personal identity. *Int. J. Creativity Probl. Solving* 28, 45–57.
- Karwowski, M., Lebuda, I., Wiśniewska, E., and Gralewski, J. (2013). Big Five personality traits as the predictors of creative self-efficacy and creative personal identity: does gender matter? *J. Creat. Behav.* 47, 215–232. doi: 10.1002/jocb.32
- Kaufman, J. C. (2006). Self-reported differences in creativity by ethnicity and gender. *Appl. Cogn. Psychol.* 20, 1065–1082. doi: 10.1002/acp.1255
- Łaguna, M., Bąk, W., Purc, E., Mielniczuk, E., and Oleś, P. K. (2014). Short measure of personality TIPI-P in a Polish sample. *Roczniki psychologiczne* 17, 421–437.
- Lemons, G. (2010). Bar drinks, rugas, and gay pride parades: is creative behavior a function of creative self-efficacy? *Creat. Res. J.* 22, 151–161. doi: 10.1080/10400419.2010.481502
- Liu, W., Pan, Y., Luo, X., Wang, L., and Pang, W. (2017). Active procrastination and creative ideation: the mediating role of creative self-efficacy. *Personal. Individ. Differ.* 119, 227–229. doi: 10.1016/j.paid.2017.07.033
- Mantel, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J. Am. Stat. Assoc.* 58, 690–700.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work and approved it for publication.

FUNDING

Support from the SRG Research Program of the University of Macau (grant number SRG2018-00141-FSS) was gratefully acknowledged.

- Maydeu-Olivares, A., and Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika* 71, 713–732. doi: 10.1007/s11336-005-1295-9
- Muck, P. M., Hell, B., and Gosling, S. D. (2007). Construct validation of a short five-factor model instrument. *Eur. J. Psychol. Assess.* 23, 166–175. doi: 10.1027/1015-5759.23.3.166
- Nunnally, J. C. (1978). *Psychometric Theory*. 2nd Edn. New York, NY: McGraw-Hill.
- Orlando, M., and Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Appl. Psychol. Meas.* 24, 50–64. doi: 10.1177/01466216000241003
- Orlando, M., and Thissen, D. (2003). Further investigation of the performance of $S-X^2$: an item fit index for use with dichotomous item response theory models. *Appl. Psychol. Meas.* 27, 289–298. doi: 10.1177/0146621603027004004
- Pretz, J. E., and McCollum, V. A. (2014). Self-perceptions of creativity do not always reflect actual creative performance. *Psychol. Aesthet. Creat. Arts* 8, 227–236. doi: 10.1037/a0035597
- Puente-Díaz, R., and Cavazos-Arroyo, J. (2017). The influence of creative mindsets on achievement goals, enjoyment, creative self-efficacy and performance among business students. *Think. Skills Creat.* 24, 1–11. doi: 10.1016/j.tsc.2017.02.007
- Qiang, R., Han, Q., Guo, Y., Bai, J., and Karwowski, M. (2020). Critical thinking disposition and scientific creativity: the mediating role of creative self-efficacy. *J. Creat. Behav.* 54, 90–99. doi: 10.1002/jocb.347
- Reiter-Palmon, R., Robinson-Morrall, E. J., Kaufman, J. C., and Santo, J. B. (2012). Evaluation of self-perceptions of creativity: is it a useful criterion? *Creat. Res. J.* 24, 107–114. doi: 10.1080/10400419.2012.676980
- Roberts, J. S. (2008). Modified likelihood-based item fit statistics for the generalized graded unfolding model. *Appl. Psychol. Meas.* 32, 407–423. doi: 10.1177/0146621607301278
- Romero, E., Villar, P., Gómez-Fraguela, J. A., and López-Romero, L. (2012). Measuring personality traits with ultra-short scales: a study of the ten item personality inventory (TIPI) in a Spanish sample. *Personal. Individ. Differ.* 53, 289–293. doi: 10.1016/j.paid.2012.03.035
- Royston, R., and Reiter-Palmon, R. (2019). Creative self-efficacy as mediator between creative mindsets and creative problem-solving. *J. Creat. Behav.* 53, 472–481. doi: 10.1002/jocb.226
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psycho. Monogr.* 17:34.
- Samejima, F. (1997). “Graded response model,” in *Handbook of Modern Item Response Theory*. eds. W. van der Linden and R. K. Hambleton (New York, NY: Springer), 85–100.
- Schack, G. D. (1989). Self-efficacy as a mediator in the creative productivity of gifted children. *J. Educ. Gifted* 12, 231–249. doi: 10.1177/016235328901200306
- Shaw, A., Elizondo, E., and Wadlington, P. L. (2020). Reasoning, fast and slow: how noncognitive factors may alter the ability-speed relationship. *Intelligence* 83:101490. doi: 10.1016/j.intell.2020.101490
- Shaw, A., Liu, O. L., Gu, L., Kardanova, E., Chirikov, I., Li, G., et al. (2020). Thinking critically about critical thinking: validating the Russian HEIghten® critical thinking assessment. *Stud. High. Educ.* 45, 1933–1948. doi: 10.1080/03075079.2019.1672640
- Silvia, P. J., Kaufman, J. C., and Pretz, J. E. (2009). Is creativity domain-specific? Latent class models of creative accomplishments and creative self-descriptions. *Psychol. Aesthet. Creat. Arts* 3, 139–148. doi: 10.1037/a0014940
- Sireci, S. G., and Rios, J. A. (2013). Decisions that make a difference in detecting differential item functioning. *Educ. Res. Eval.* 19, 170–187. doi: 10.1080/13803611.2013.767621
- Steinberg, L., and Thissen, D. (1995). “Item response theory in personality research,” in *Personality Research, Methods, and Theory: A Festschrift Honoring Donald W. Fiske*. eds. P. E. Shrout and S. T. Fiske (Hillsdale, NJ: Erlbaum), 161–181.
- Tierney, P., and Farmer, S. M. (2002). Creative self-efficacy: its potential antecedents and relationship to creative performance. *Acad. Manag. J.* 45, 1137–1148. doi: 10.5465/3069429
- Tierney, P., and Farmer, S. M. (2011). Creative self-efficacy development and creative performance over time. *J. Appl. Psychol.* 96, 277–293. doi: 10.1037/a0020952
- Zieky, M. (1993). “DIF statistics in test development,” in *Differential Item Functioning*. eds. P. W. Holland and H. Wainer (Hillsdale, NJ: Erlbaum), 337–347.

Conflict of Interest: NM was employed by company Codility Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Shaw, Kapnek and Morelli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.