Check for updates

Feature Article

Applying Automated Originality Scoring to the Verbal Form of Torrance Tests of Creative Thinking

Gifted Child Quarterly
1–15
© 2021 National Association for Gifted Children
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00169862211061874
journals.sagepub.com/home/gcq
SAGE

Selcuk Acar[1], Kelly Berthiaume[1], Katalin Grajzel[2], Denis Dumas[2], Charles "Tedd" Flemister[1], and Peter Organisciak[2]

Abstract

In this study, we applied different text-mining methods to the originality scoring of the Unusual Uses Test (UUT) and Just Suppose Test (JST) from the Torrance Tests of Creative Thinking (TTCT)–Verbal. Responses from 102 and 123 participants who completed Form A and Form B, respectively, were scored using three different text-mining methods. The validity of these scoring methods was tested against TTCT's manual-based scoring and a subjective snapshot scoring method. Results indicated that text-mining systems are applicable to both UUT and JST items across both forms and students' performance on those items can predict total originality and creativity scores across all six tasks in the TTCT-Verbal. Comparatively, the text-mining methods worked better for UUT than JST. Of the three text-mining models we tested, the Global Vectors for Word Representation (GLoVe) model produced the most reliable and valid scores. These findings indicate that creativity assessment can be done quickly and at a lower cost using text-mining approaches.

Keywords

creativity assessment, gifted identification, originality scoring, scale validation, text-mining, Torrance Tests of Creative Thinking–Verbal

Creativity has been viewed as a major component or outcome in many models (Gagné, 2005; Renzulli, 2005; Renzulli & Reis, 2014; Sternberg, 2003; Subotnik & Jarvin, 2005) and definitions of giftedness (Marland, 1972; Rinn et al., 2020). According to the 2018–2019 State of the States in Gifted Education Report (Rinn et al., 2020), creativity was explicitly mentioned in 31 states' definition of giftedness in the United States. Beyond gifted education, creativity is a relevant, in-demand skill for the workforce and innovation-centered economy (Frey & Osborne, 2017; Lichtenberg et al., 2008; Petrone, 2019). In the revised form of Bloom's taxonomy (Krathwohl, 2002), "creating" is the highest level of thinking. These practical and theoretical accounts underline the need for further study of creativity and educational programming to foster it for all students, consistent with contemporary educational frameworks (e.g., Partnership for 21st Century Skills, 2006).

The field of gifted education has given particular attention to creativity. Giftedness without creativity may imply raising information consumers rather than information producers, innovators, groundbreakers, and change-makers (Wiley & Voss, 1996). From a conceptual and psychometric standpoint, decades of research have indicated that intelligence and creativity have smaller relationships to each other at higher levels of intelligence, which is known as the threshold hypothesis (Dumas, 2018; Guilford, 1967; Jauk et al., 2013; Karwowski et al., 2016), although see Weiss and colleagues (2020) for recent counter-evidence to this hypothesis. Thus, assessment of intelligence alone does not cover creative giftedness (Castejón et al., 2016; Kim et al., 2013; Zenasni et al., 2016). Furthermore, failure to assess creativity in gifted identification may lead to favoring and overidentification of *schoolhouse giftedness* as a result of excluding students who think differently, challenge the status quo, question authority, and go beyond the accepted paradigms (Kaufman et al., 2012; Renzulli, 2005). To avoid this potential pitfall, identification of gifted students includes creativity assessment in certain states (e.g., Georgia, Arkansas) and districts (Krisel & Cowan, 1997; Peters et al., 2020; Rinn et al., 2020).

[1]University of North Texas, Denton, TX, USA
[2]University of Denver, CO, USA

Corresponding Author:
Selcuk Acar, Department of Educational Psychology, University of North Texas, Matthews Hall 304E, 1300 W. Highland Street, Denton, TX 76201, USA.
Email: selcuk.acar@unt.edu

Gifted Identification

Current literature on gifted identification urges educators to apply multiple criteria to identify students for gifted programs (Acar et al., 2016; McBee et al., 2016), but how those multiple criteria are used also is of great importance (Lee et al., 2020; Peters et al., 2020). Given the representation discrepancy in gifted identification by race (Grissom & Redding, 2016), English-language learner status (Mun et al., 2020) and socioeconomic status (Grissom et al., 2019; Peters et al., 2019), strategies such as universal screening and universal consideration (McBee et al., 2016), the use of local norms (Peters et al., 2019), and the use of alternative assessment (Silverman & Gilman, 2020) seem to have the potential to improve the representational fairness of the identification process.

Creativity is one such alternative assessment approach that may improve underrepresentation of children who are from low socioeconomic status families, English-language learners, or racial/ethnic minorities (Kaufman et al., 2012; Luria et al., 2016; Matthews, 2015). The use of a creativity assessment does not need much extra justification on the basis of conceptualizations of giftedness because most theoretical conceptualizations of giftedness include creativity in their theorizing (Gagné, 2005; Renzulli, 2005; Sternberg, 2003; Subotnik & Jarvin, 2005). However, one could argue that there is no point of including a creativity measure in identification if programs and services for gifted students do not explicitly aim to foster creativity. Indeed, discarding the measures of creativity could appear to provide a better alignment between the programs and identification, but this leads to a more concerning misalignment, which is between the services/programs and the field's theoretical conceptualization of giftedness. Thus, although creativity is key to giftedness and including creativity measures in gifted identification is consistent with many conceptualizations of giftedness, this is most critical when the programs and services offered in the gifted programs target or involve creativity and creative thinking skills. Although there are a number of alternatives for capturing creativity, such as consensual assessment technique and checklists or ratings scales by teachers, peers, or parents (see Kaufman et al., 2012), the Torrance Tests of Creative Thinking (TTCT; Torrance, 2008) has been the most popular tool used for gifted identification (Hunsaker & Callahan, 1995).

TTCT

The TTCT builds on the concept of divergent thinking, which originated from Guilford's (1967) Structure of Intellect (SOI) model. Divergent thinking tasks are open-ended and allow responding to the tasks in various ways (i.e., "In what ways can you use a brick?" "List things that move on wheels" "How are a pencil and pajamas alike?" "What does this figure look like?"). Lack of a set of defined acceptable or correct responses has been a defining characteristic of divergent thinking tasks (and of most challenges requiring creative thinking), but this characteristic also makes the scoring of divergent thinking tasks time-consuming and laborious. The traditional scoring method involves the indices of fluency (the number of relevant, meaningful responses), flexibility (the number of distinct clusters or categories in the produced responses), originality (statistically infrequent responses), and elaboration (amount of detail and elegance of the responses). Among those, originality is the most essential component of creativity (Acar et al., 2016; Diedrich et al., 2015), which is the focus of the present study.

There are various versions of the TTCT such as Figural, Verbal, and Abbreviated. The present study focuses on scoring of originality in the verbal form of the TTCT. TTCT-Verbal, the focus of the current study, has two parallel forms—Form A and Form B—both of which consist of six activities: asking, guessing causes, guessing consequences, toy improvement, unusual uses, unusual questions, and just suppose test. Each activity is scored on three indices: fluency, flexibility, and originality. The norms are available for both forms from age 6 and above. Said-Metwali et al. (2021) conducted a factor analysis of the Arabic version of the TTCT-Verbal that yielded a second-order factor structure where each activity formed its own factor under a general divergent thinking factor. They found poor discriminant validity among the three indices, which is a likely consequence of a fluency confound in the originality and flexibility scores (Forthmann et al., 2020). Said-Metwali et al. (2021) also reported evidence supportive of measurement invariance by gender and academic major that are partly consistent with Krumm et al.'s (2016) results that, again, reported six factors, where each activity forms its own factor. However, Krumm et al. did not explore a general higher-order factor.

These factor analytic studies used all three indices (i.e., fluency, flexibility, and originality) and, due to fluency confound, it could be argued that scoring the tasks for all three may not be always necessary due to high overlap among the three scores despite their theoretical distinctness and importance. This high overlap also explains why each activity forms its own factor rather than fluency, flexibility, and originality making up three factors. To make sense of it, a closer look at how originality is scored could help.

Originality scoring in TTCT-Verbal is based on zero-originality lists, which supply the common responses for each activity. The scorers use the zero-originality lists to determine whether a response deserves a point for originality. For example, "animal house" or "throw away" is provided in the zero-originality list for Activity 5, which asks to list different uses for a cardboard box. Hence, those responses would get zero points, whereas a response such as "shield" would get 1 because it is not on the zero-originality lists, yet it makes sense. The responses that are on the zero-originality lists do not get any originality points, whereas those that are relevant

and not on the zero-originality lists get 1 point for originality. Therefore, as more responses are produced by a participant, originality also increases proportionally with fluency. Alternative scoring procedures may prove useful and potentially could avoid any undue overlap among originality and fluency.

Currently, TTCT-Verbal is scored by trained, certified raters who compute the scores manually using the guidelines in the TTCT manual. This makes the scoring process quite time-consuming and costly especially when a large number of people take the tests, which would be the case with universal screening and consideration. As a matter of fact, the presence of zero-originality lists (and the list of categories for flexibility scoring) is more practical than developing sample-based zero-originality lists, which is how other divergent thinking tasks are often scored. Still, the cost of scoring is a huge burden on school districts and limits the usability of the TTCT-Verbal (Glover & Albers, 2007). Consequently, it may deter many school districts from using such standardized tests and rely instead on other forms of assessment such as teacher referral, or resort to "home-made" creativity measures that lack solid psychometric evidence. Ideally, more of schools' resources should be used for programming and differentiation than for assessment and identification. Therefore, novel methods and approaches are necessary to obtain psychometrically robust and cost-effective measurement of creativity. A recent approach, known as semantic distance, allows for the rapid and automated scoring of divergent thinking tasks (Acar et al., 2020; Dumas & Dunbar, 2014) and may be applicable to at least some activities on the TTCT-Verbal.

Automated Scoring of Divergent Thinking Tasks

The use of semantic networks as a method for scoring originality has recently gained attention in creativity research (Acar & Runco, 2014; Dumas & Dunbar, 2014; Forster & Dunbar, 2009). In one such investigation, Acar and Runco (2014) used words and concepts obtained from three different associative networks (WordNet, Word Association Network, and IdeaFisher) as if they were socially established zero-originality lists; the mention of the words and concepts from these networks in the responses imply making a close association and thereby lower originality. The WordNet is a lexical network with words organized by relations on several levels such as hyponymy (subtype relations such as "horse" is a subtype of "animal") and meronymy (supertype or part-whole relations such as "arm" is a part of "body"). The Word Association Network is a website that compiles vocabulary based on its use in classical and contemporary literature. Finally, the IdeaFisher is a system built on word association for the purpose of priming the retrieval of words with semantically close other words. Although these three networks vary in size, Acar and Runco (2014) scored the responses as original or unoriginal based on the utilization of the words and concepts provided by each of the networks, respectively. These researchers reported that originality scores obtained in this way had a significant correlation with the attitudes and values toward creativity. They also found more semantically distant responses in the second half than the first half of the response list produced by the participants. Known as the order effect (Beaty & Silvia, 2012; Runco, 1986), this has been one of the most replicable findings on originality and its applicability to semantics-based originality adds to the validity of this scoring method.

Latent Semantic Analysis

Although Acar and Runco (2014) demonstrated the usefulness of word association networks, their method was not completely automated. Thus, to improve the efficiency and objectivity of divergent thinking measures without the use of human raters, creativity researchers have used some specific text-mining methods such as latent semantic analysis (LSA) through which semantic distances are calculated (Acar & Runco, 2019; Dumas & Dunbar, 2014; Forster & Dunbar, 2009; Hass, 2017b; Heinen & Johnson, 2018; Prabhakaran et al., 2014). LSA is a highly effective quantitative text analytics method that estimates the latent similarity between words or phrases by using a large corpus of text representing the meaning of words and phrases specific to the context of interest (e.g., English language, children; Dumas et al., 2020; Prabhakaran et al., 2014). LSA analyzes the text-corpus to build a matrix in which vectors represent the co-occurrences of words (Chen et al., 2011; Dumas & Dunbar, 2014). Semantic distance is then calculated from the cosine of the angle between the two vectors and subtracted from 1 to produce an originality score (Dumas & Dunbar, 2014; Dumas et al., 2020; Heinen & Johnson, 2018). A greater semantic distance is indicative of less similarity between the divergent thinking task prompt and response and therefore a greater originality score.

One potential factor that is likely to influence these originality scores is the corpus used to model the word co-occurrences. The magnitude (e.g., number of words) and scraping method (e.g., source of the words) of the corpus might influence the outcomes. The topical domain of the corpora may also affect the resulting model. In the past, creativity research has typically employed models trained on corpora of general language (Dumas et al., 2020) or of language extracted from educational texts (Dumas & Dunbar, 2016; Landauer & Dumais, 1997). It is possible that a corpus tuned to child-specific texts (e.g., Wild et al., 2013) would be more appropriate for working with measurement for children, although, at this point, the strength of this effect is not known.

Several studies have demonstrated LSA is a more efficient, reliable, and valid means of scoring originality on divergent thinking tasks compared with traditional scoring methods (Acar & Runco, 2019; Dumas & Dunbar, 2014;

Forster & Dunbar, 2009; Heinen & Johnson, 2018). Early work by Dunbar and colleagues supported that LSA scores were stronger predictors of human-rated originality compared with fluency- and elaboration-based assessments (Forster & Dunbar, 2009). In addition, Forster and Dunbar (2009) suggested that LSA originality scoring can discriminate between creative use and common use prompts for Alternative Uses Tests (Guilford, 1967). Later, Dumas and Dunbar (2014) examined the relationship between originality and fluency using LSA originality latent scores on the Alternate Uses Test. Findings from this study revealed that LSA-based originality scores can have greater reliability compared with fluency scores. In addition, Dumas and Dunbar (2014) established discriminant validity between LSA originality scores and fluency scores and proposed that using LSA for divergent thinking scoring reduces measurement error, giving creativity researchers stronger predictive validity to determine originality scores from divergent thinking tasks. Results from Dumas and Dunbar's (2014) study bolstered the claims that operationalizing originality as semantic distance, using LSA techniques, is a psychometrically beneficial, productive, and objective method for originality measurement.

Several other creativity researchers have provided strong support for discriminant (Dumas & Dunbar, 2016; Dumas & Runco, 2018; Gray et al., 2019; Prabhakaran et al., 2014), convergent (Forthmann, Oyebade, et al., 2019; Hass, 2017b; Heinen & Johnson, 2018), predictive (Dumas & Strickland, 2018; Dumas et al., 2020), and concurrent (Beketayev & Runco, 2016; Dumas et al., 2020) validity for LSA Originality scoring. Hass (2017a) provided evidence of validity for using LSA to calculate local similarity (semantic relationship between adjacent responses) and global similarity (semantic proximity between prompts and each response) on the Alternate Uses Test. Hass found a significant, negative association between human-judged creativity ratings and both local and global similarity. Hass (2017a) also reported that similar responses were generated in less time compared with dissimilar responses. In a different study, Hass (2017b) used LSA to examine changes in semantic distance across successive response iterations during divergent thinking tasks on the Alternate Uses Test. Results of this study demonstrated parallel results to the past research in that participants with higher fluid intelligence generated responses with greater semantic distance (i.e., greater originality), despite the linear increase in semantic distance across response iterations. Gray et al. (2019) used LSA to calculate forward flow or the semantic distance between responses over time across multiple studies. Similar to findings from Hass (2017a), these studies supported that forward flow was positively related to and predicted creativity across diverse domains and divergent thinking tasks, and demonstrated discriminant validity between fluency and originality. Later, Forthmann, Oyebade, et al. (2019) investigated whether elaboration, or the number of words used in a response, confounds LSA-derived semantic distance. Elaboration was, indeed, associated with declines in semantic distance scores even with stop word (e.g., the, or, and) removal and this is a limitation of LSA-based originality scores.

Since the initial findings on LSA-based originality scoring, many creativity researchers have further demonstrated the essential and fruitful applications of using semantic distance to operationalize originality. The objectivity of LSA-derived originality scores has allowed researchers to examine long-standing research questions in the field of creativity and to understand and estimate the relationships between creativity and other psychological phenomena with superior precision (Dumas et al., 2020; Dumas & Strickland, 2018). For instance, Dumas (2018) found LSA-derived originality scores on the Alternate Uses Test were associated with reasoning ability, but only among participants with originality scores at or below the median. This finding supported the threshold hypothesis, which describes intelligence as a necessary but not sufficient condition for creativity. However, these findings somewhat differed from the traditional threshold hypothesis in terms of the threshold variable, creativity (i.e., originality) rather than cognitive (i.e., relational reasoning), and the location of the threshold, the median of originality rather than the traditional intelligence quotient level of 120 (Dumas, 2018).

LSA methods have also allowed creativity researchers to examine the effects of specific instructions or cue conditions on originality scores and to make inferences regarding their relation to divergent thinking. For example, Prabhakaran and colleagues (2014) found that semantic distance of verb-word pairs, which was measured by LSA, was larger when the respondents were cued to be creative while generating verbs for nouns. Likewise, Heinen and Johnson (2018) also showed that creativity of the responses as measured by semantic distance is influenced by the type of explicit instructions presented. Both semantic distance and subjective creativity ratings of the responses were higher when instructions emphasized novelty rather than appropriateness. In another study, Dumas and Dunbar (2016) found that participants' LSA-based originality scores were significantly lower when divergent thinking tasks were framed with stereotypes related to creative ("an eccentric poet") and noncreative ("a rigid librarian") persons than under the control condition where no such stereotype was presented. They concluded that the use of stereotypes in the instructions may lead to inhibition and suppress the divergent thinking outcomes. Together, these findings illustrate sensitivity of LSA-based semantic distance metrics to explicit instructions and cues, which parallels extant research on the impact of explicit instructions on divergent thinking tasks when traditional scoring methods were used (Acar et al., 2020; Said-Metwaly et al., 2020).

Moving Beyond the Latent Semantic Analysis

Importantly, there are alternative text-mining methods to the LSA. Dumas et al. (2020) recently conducted a psychometric

evaluation to assess the reliability and criterion validity of four common text-mining systems (i.e., Touchstone Applied Science Associates [TASA] LSA, English 100k LSA, Global Vectors for Word Representation [GloVe] 840B, and Word2Vec; see Dumas et al., 2020, for detailed descriptions of text-mining systems) used to calculate semantic distance on the Alternate Uses Test compared with human-completed ratings. Although this study found that human-rated originality scores demonstrated the highest reliability and better discriminant validity from fluency and elaboration, text-mining systems, particularly the GloVe 840B system, are capable of generating highly reliable and valid originality scores in a more economical and automated manner. Furthermore, there are open-access platforms (see https://openscoring.du.edu/ for one example) to utilize those systems, which presents a unique opportunity for gifted identification efforts to score divergent thinking tasks rapidly, consistently, and at no financial cost.

The Present Study

Given the low-cost and quick scoring provided by the automated scoring platforms, the implication of automated scoring of divergent thinking tasks is obvious for gifted identification, especially for school districts that aim to include creativity assessment in gifted identification and adopt a universal screening approach. Typically, automated scoring is applied to Alternate Uses tasks (Dumas & Dunbar, 2014; Hass, 2015, 2017b) and its applicability to other types of divergent thinking tasks, such as Consequences, is rather new. LaVoie et al. (2020) applied latent semantic analysis to the Consequences task and found a high level of agreement between LSA-based scores and human ratings. Text-mining methods, including, but not limited to, LSA, have yet to be tested with specific activities of the TTCT. In the present study, we examine the applicability of text-mining based originality scoring procedures to two different types of tasks: Unusual Uses Test (UUT) and Just Suppose Test (JST), analogous to Alternate Uses and Consequences, respectively. In the present study, we assess their applicability through validation with alternate scoring methods (i.e., snapshot scoring and TTCT-manual-based scoring). Furthermore, we do so by comparing various text-mining systems (e.g., GLoVe, TASA, EN 100k) that are easy to access and use. If automated scoring is applicable to both of those tasks, it is likely that they could serve as a short form of TTCT-Verbal and may be used in universal screening for gifted identification. Ultimately, school districts could save on testing and identification and allocate more of their resources to programming and talent development.

In the present investigation, we address the following research questions:

Research Question 1: Are automated originality scores of UUT and JST related to snapshot originality scores?

Research Question 2: Are automated originality scores of UUT and JST related to TTCT-manual-based originality?

Research Question 3: Which automated scoring system (GLoVe, TASA, EN 100k) provides better predictive power?

Method

Participants

The study sample consisted of 225 participants who completed either Form A ($n = 102$) or Form B ($n = 123$) of the TTCT-Verbal. Among those for whom demographic information was available, 60 (26.7%) were male, 110 (48.9%) were female, and 55 (24.4%) did not report, with an average age of 18.20 ($SD = 1.31$) years. Participants were a largely general (i.e., nonselective) group of undergraduate students who completed the TTCT as part of an intervention program that aimed to facilitate adaptation to college life.

Instruments

The only instrument administered was the TTCT-Verbal (Torrance, 2008), which consisted of six activities. The responses to two of those activities were used in the present study: UUT and JST. The UUT presents an everyday object and asks participants to write down different ways to use this object. The UUT dates back to Guilford's (1967) battery and is the most commonly used type of divergent thinking task (Runco et al., 2016). The everyday objects used in TTCT-Verbal are cardboard boxes and tin cans in Form A and Form B, respectively. The JST presents a hypothetical situation for participants to think of possible consequences if it were to happen, similar to the well-known Consequences task (Christensen et al., 1958). In TTCT-Verbal, the hypothetical situations for JST are: " . . . the clouds had strings attached to them . . . What would happen?" in Form A (JST-Cloud Strings) and " . . . all we could see of people would be their feet . . . What would happen?" in Form B (JST-See Feet). As typical in most divergent thinking tasks, both UUT and JST instructed the participants to generate as many responses as possible. The UUT additionally emphasized the generation of a variety of responses that other people would not think of. The participants were given 5 min to complete each activity.

Procedures

We measured originality of the produced responses in different ways. These fall under three major categories: (a) snapshot, (b) TTCT-manual-based, and (c) text-mining-based.

Snapshot Originality. After participants' written responses were transcribed, two raters (i.e., two trained graduate students) independently rated the responses of each participant by using

the ideational pools scoring method (recently renamed as snapshot scoring; Runco & Mraz, 1992; Silvia et al., 2009). The raters were trained through example responses that had quintessentially high versus low originality and instructions on how to assign the rating scores based on a normal distribution. After some preliminary work on a subset of responses, we made clarifications to the training based on the raters' questions. In this subjective scoring method, the raters evaluated each participant's response set generated for a divergent thinking prompt (rather than individual responses) in terms of their originality, cleverness, and remoteness on a 7-point scale (1 = *least original*, 7 = *most original*). The raters were trained to rate the response set by approximating the variation to a normal distribution, with most response sets falling in the middle of the 7-point scale. The intraclass correlation (ICC) (1,2) was .76 (95% confidence interval [CI] = [.65, .84]) and .78 (95% CI = [.69, .85]) for the UUT and .71 (95% CI = [.61, .77]) and .69 (95% CI = [.57, .81]) for JST items. The subjective ratings of the two raters were averaged to form the "Snapshot originality" variable.

TTCT-Manual-Based Originality. The manual-based scoring of TTCT-Verbal was conducted by the Scholastic Testing Services, the publisher of the test. According to the test manual (Torrance, 2008), originality of each relevant responses is scored on a binary metric. When a response is provided on the zero-originality lists, which is provided as part of the scoring manual, it gets no points. And, if the relevant responses are not provided on the zero-originality lists, then each of such instances is scored as one. The count of original responses (summed total) is used as the originality score of each participant.

Text-Mining-Based Originality. The automated scoring of originality was obtained from the Open Creativity Scoring (https://openscoring.du.edu/) freeware platform (Dumas et al., 2020; Organisciak & Dumas, 2020). This platform provides originality scores for verbal divergent thinking stimuli and has been previously validated for the UUT task, but at the time of this research, had never been applied to the JST task (Dumas et al., 2020). The Open Creativity Scoring freeware can instantly generate originality scores for each participant's responses, thereby saving significant time and resources. Compared with the laborious task of scoring each answer by hand, these text-mining systems allow for efficiency and accuracy (Dumas & Dunbar, 2014; Forthmann, Oyebade, et al., 2019). Text-mining models are presented with a large amount of textual data (termed a corpus) to establish patterns of words used in everyday language. Dimension reduction techniques, such as latent semantic analysis (LSA), are then used to establish categories for similar words (synonyms). As a result, words that are similar are located closer in semantic space, whereas more unique answers earn higher scores due to their larger semantic distance from the prompt (Dumas et al., 2020).

There are many freely available text-mining systems available on the web today. Although they all use similar methods, they also differ in several important ways, including in the kind and extent of the text library or corpora they are based on, the method of training, the specific statistical models, the corrections for real-world scenarios such as words with multiple meanings, and the correction for commonly used function words (Dumas et al., 2020). The Open Creativity Scoring platform used in this study allows stoplisting as an option to filter out common function words. Stoplisting permits the omission of words such as "the," "to," and "make." These words are used to connect grammatical parts of the sentence and contribute minimally to meaning (Dumas et al., 2020; Fox, 1989).

We used three different text-mining systems to compare originality scores of participants, all of which are available for free on the Open Creativity Scoring platform. *TASA* (Touchstone Applied Science Association LSA) by Landauer and Dumais (1997) is the most widely used system in divergent thinking research (Dumas & Dunbar, 2014; Forster & Dunbar, 2009; Forthmann, Oyebade, et al., 2019). It has been trained on a corpus of 37,651 educational texts to mirror the reading level of an undergraduate student. *EN 100k* by Günther et al. (2015) includes 100,000 of the most frequently used English words. It uses LSA, similarly to TASA, but includes an extended corpus trained on mostly British web-based resources (e.g., Wikipedia, ukWac). *GLoVe* (Global Vectors for Word Representation 840B) was created by the Stanford natural-language-processing laboratory (Pennington et al., 2014). It was trained on a corpus of 840 billion words, including web-based documents. Compared with TASA and EN 100k, GLoVe uses a probabilistic modeling framework and examines co-occurrence of words in a small space, rather than in full-text documents.

Responses are considered more original to the extent they are semantically distant from the presented prompt (e.g., cardboard boxes). We calculated total originality and the average originality by dividing total originality by fluency for each of the three scores per activity. All in all, we had the following originality scores for each activity: (a) *Snapshot* originality obtained from two raters' subjective evaluations using total ideational output, (b) the *TTCT-manual-based* originality scores, (c) *Averaged GLoVe-based* originality, (d) *Averaged TASA-based* originality, (e) Averaged EN-100k-based originality, (f) *Total GLoVe-based* originality, (g) *Total TASA-based* originality, and (h) *Total EN 100k-based* originality. The present study reports the correlations among these indices and tests the relationships among three different automated text-mining scores and two different criteria measures, namely, Snapshot originality and the TTCT-manual-based originality scores. Using Cohen's (1988) guidelines, correlations above .30 (medium to high) are considered useful and psychometrically meaningful.

Table 1. Descriptive Statistics of Originality Scores for the UUT and JST.

	UUT				JST			
	Cardboard boxes (n = 102)		Tin cans (n = 120)		Cloud strings (n = 102)		See feet (n = 121)	
Originality scores	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1. Snapshot	4.49	1.54	4.40	1.59	4.86	1.42	4.28	1.42
2. Manual-Based	10.20	5.90	8.66	5.77	8.19	5.83	8.42	5.00
3. GLoVe Average	0.68	0.06	0.70	0.07	0.68	0.06	0.75	0.06
4. TASA Average	0.87	0.07	0.89	0.09	0.87	0.07	0.94	0.06
5. EN 100k Average	0.54	0.05	0.55	0.07	0.54	0.05	0.64	0.06
6. GLoVe Total	7.05	4.24	7.19	4.60	7.05	4.24	7.26	4.09
7. TASA Total	9.04	5.47	9.20	5.83	9.04	5.47	9.15	5.09
8. EN 100k Total	5.63	3.44	5.67	3.56	5.63	3.44	6.25	3.56

Note. UUT = Unusual Uses Test; JST = Just Suppose Test; GLoVe = Global Vectors for Word Representation; TASA = Touchstone Applied Science Association.

Table 2. Correlations Among Various Originality Scoring Methods of Unusual Uses Activity of TTCT-Verbal—Form A (Lower Diagonal) and Form B (Upper Diagonal).

Originality scoring methods	1	2	3	4	5	6	7	8
1. Snapshot		.679**	.324**	.334**	.110	.702**	.699**	.690**
2. Manual-Based	.623**		.216*	.198*	.079	.955**	.953**	.946**
3. GLoVe Average	.333**	.200*		.888**	.856**	.332**	.311**	.348**
4. TASA Average	.235*	.116	.760**		.739**	.327**	.330**	.336**
5. EN 100k Average	.155	.029	.868**	.718**		.175	.153	.225*
6. GLoVe Total	.622**	.956**	.347**	.260**	.193		.998**	.996**
7. TASA Total	.613**	.951**	.304**	.293**	.165	.993**		.992**
8. EN 100k Total	.613**	.940**	.359**	.282**	.248*	.995**	.989**	

Note. N = 102 Form A (Lower Diagonal). N = 120 Form B (Upper Diagonal). TTCT = Torrance Tests of Creative Thinking; GLoVe = Global Vectors for Word Representation; TASA = Touchstone Applied Science Association.
*$p < .05$. **$p < .01$.

Results

Table 1 presents the descriptive statistics for each originality score across four divergent thinking tasks from the TTCT-Verbal.

The UUTs

The correlations among originality scores were analyzed and reported for each activity. Table 2 presents the correlations for the UUT items. We reported the correlations with total and average automated (text-mining-based) originality scores.

Analyses With Total Originality Scores. When total automated scores for UUT-Cardboard Boxes (Form A) are considered, analyses indicated that all three automated scores had a statistically significant and positive correlation with both Snapshot (rs $> .61$, $p < .01$) and TTCT-manual-based scores (rs $> .94$, $p < .01$). There was a similar pattern in the analyses with total automated scores of UUT-Tin Cans (Form B) such that the correlation of total automated scores was significant for both Snapshot originality (rs $> .69$, $p < .01$) and TTCT-manual-based originality scores (rs $> .94$, $p < .01$).

To examine differences among automated scores, we compared the correlations of three automated originality scores (i.e., GloVe, TASA, EN 100k) with Snapshot and TTCT-manual-based originality scores. In the Cardboard Boxes task, the correlations with Snapshot originality did not significantly differ by the type of automated scores based on Fisher's z-test, whereas GloVe-based originality scores had a significantly higher correlation with TTCT-manual-based originality scores than did EN 100k ($z = 3.49$, $p < .01$). The difference of the correlations was not significant between GloVe and TASA ($z = 0.96$, $p = .17$). For the UUT-Tin Cans, GLoVe had a significantly higher correlation with Snapshot originality than EN 100k ($z = 2.02$, $p = .02$), whereas comparisons involving TASA were not significant (zs = 0.72 and −1.07, ps = .23 and .14). With TTCT-manual-based scores, GloVe and TASA had significantly higher correlations (zs =

Table 3. Correlations Among Various Originality Scoring Methods of Just Suppose Activity of TTCT-Verbal—Form A (Lower Diagonal) and Form B (Upper Diagonal).

Originality scoring methods	1	2	3	4	5	6	7	8
1. Snapshot		.551**	.319**	.170	.119	.545**	.531**	.526**
2. Manual-Based	.511**		.146	.145	.217*	.962**	.961**	.956**
3. GLoVe Average	.045	.095		.767**	.819**	.318**	.290**	.320**
4. TASA Average	.041	.109	.843**		.724**	.206*	.218*	.219*
5. EN 100k Average	.067	.102	.904**	.866**		.293**	.279**	.337**
6. GLoVe Total	.511**	.933**	.216*	.236*	.220*		.998**	.996**
7. TASA Total	.510**	.931**	.197*	.245*	.213*	.998**		.995**
8. EN 100k Total	.506**	.926**	.233*	.264**	.262**	.997**	.997**	

Note. $N = 102$ Form A (Lower Diagonal). $N = 120$ Form B (Upper Diagonal). TTCT = Torrance Tests of Creative Thinking; GLoVe = Global Vectors for Word Representation; TASA = Touchstone Applied Science Association.
*$p < .05$. **$p < .01$.

3.55 and 1.95, ps $< .01$ and $= .026$, respectively) than EN 100k. The difference of the correlations between TASA and GloVe originality scores was not significant ($z = 1.147$, $p = .126$).

Analyses With Average Originality Scores. For UUT-Cardboard boxes, Snapshot originality was significantly related to GLoVe-based and TASA-based average originality scores (rs $= .33$ and .24, ps $< .05$ and .01, respectively) but not with EN 100k-based scores ($r = .16$, $p = .12$). With TTCT-manual-based originality scores, only the GLoVe-based originality was significantly related ($r = .20$, $p = .044$). For UUT-Tin Cans, Snapshot originality was again significantly related to both GLoVe-based and TASA-based originality (rs $= .33$ and .33, ps $< .01$), but not with EN 100k (rs $= .11$, $p = .23$). TTCT-Manual-based originality was also significantly related to both GLoVe-based and TASA-based originality (rs $= .22$ and .20, ps $= .018$ and .03), but not with EN 100k (rs $= .08$, $p = .39$).

The JST

Table 3 presents the correlations among the originality scores for both JST tasks.

Analyses With Total Originality Scores. Analyses with JST-Cloud Strings showed that Snapshot originality was significantly related to GLoVe, TASA, and EN 100k-based originality scores (rs $> .50$, ps $< .01$). The correlations of those three automated originality scores were significantly related to TTCT-manual-based scores (rs $> .92$, ps $< .01$). With JST-See Feet, the correlations of the automated scores were again significantly related to both Snapshot (rs $> .52$, ps $< .01$) and TTCT-manual-based scores (rs $> .95$, ps $< .01$).

We then compared the correlations from the JST-Cloud Strings across three automated originality scores. The correlations of the automated scores with Snapshot originality did not vary significantly across GLoVe, TASA, and EN 100k (zs > 0.18, $p > .228$). For the TTCT-manual-based scores, the correlations of both GLoVe and TASA were significantly higher than EN 100k (zs $= 2.45$ and 2.13, ps $= .007$ and .017) and the difference of the correlations between TASA and GLoVe was not significant ($z = 0.87$, $p = .23$). With JST-See Feet, both GLoVe and TASA had significantly higher correlations with Snapshot originality (zs $= 2.84$ and 2.72, ps $= .003$ and .002) than EN 100k and the difference between GLoVe and TASA was not significant (zs $= 0.637$, $p = .26$). Both GLoVe and TASA had significantly higher correlations (zs $= 2.50$ and 1.93, ps $= .005$ and .027) with the TTCT-manual-based scores than EN 100k and the difference between GLoVe and TASA was not significant ($z = 0.62$, ps $= .27$).

Analyses With Average Originality Scores. When using average originality scores across the four tasks (rather than total or summed originality scores), correlations were significant for the JST-See Feet between Snapshot originality and GLoVe-based scores ($r = .32$, $p < .01$) and TTCT-manual-based scores and EN 100k ($r = .22$, $p = .017$). The correlations for JST-Cloud Strings were not significant with average scoring.

Predicting Total Test Originality Scores With Automated Scoring

Given the practical benefits of administering fewer items, we examined the amount of variance explained by the UUT and JST items in total originality scores across all six activities in TTCT-Verbal. We regressed aggregate originality scores across all six activities on total automated UUT and JST originality scores using the forced entry method, in which both predictors were added to the model together in a single step. UUT and JST originality scores from GLoVe-total explained 55.2% of the variance, $F(2, 99) = 60.92$, $p < .001$, in Form A and 67.4% of the variance, $F(2, 117) = 120.84$,

$p < .001$, in Form B. When TASA-originality scores were used as the predictors, UUT and JST items explained 55.4% of the variance in Form A, $F(2, 99) = 61.51$, $p < .001$, and 68% in Form B, $F(2, 117) = 124.03$, $p < .001$. With EN 100k, explained variance is 28.7% for Form A, $F(2, 99) = 19.94$, $p < .001$, and 56.2% in Form B, $F(2, 117) = 75.08$, $p < .001$.

Predicting Overall Verbal Creativity With Automated Scoring

Considering the high overlap among the three indices (i.e., fluency, originality, and flexibility) and lack of discriminant validity (Said-Metwaly et al., 2020), we examined whether the automated scores from UUT and JST could predict total verbal creativity, inclusive of originality, fluency, and flexibility. Again, the forced entry method was used in the regression model in a single step. With GLoVe, we found that UUT and JST explained 67.1% of the variance, $F(2, 99) = 101.08$, $p < .001$, in Form A and 72.4% of the variance, $F(2, 117) = 153.32$, $p < .001$, in Form B. When TASA-originality was used as a predictor, the model explained 67.7% of the variance, $F(2, 99) = 103.75$, $p < .001$, in Form A and 73.1% of the variance, $F(2, 117) = 158.96$, $p < .001$, in Form B. Originality scores based on EN 100k predicted 65% of the variance, $F(2, 99) = 91.92$, $p < .001$, in Form A and 72.4% in Form B, $F(2, 117) = 153.16$, $p < .001$.

When the above analyses were replicated with TTCT-manual-based scores (replacing the automated scores as the predictors), the regression models had similar levels of predictive power in both Form A, $R^2 = .72$, $F(2, 99) = 124.71$, $p < .001$, and Form B, $R^2 = .73$, $F(2, 119) = 164.18$, $p < .001$.

Discussion

Gifted identification involves high-stakes decision-making. Therefore, the assessment tools used in the process must be psychometrically robust and cost-effective so that students can be recruited to gifted and talented programs with solid criteria without using a large proportion of financial resources. This becomes more challenging when universal screening and multiple-criteria approaches— a recommended framework for gifted identification (Peters et al., 2020)—are adopted. In the present study, we focused on creativity assessment and tested the suitability of items from two tasks from TTCT-Verbal for automated scoring of originality. Overall, our findings are solid and promising toward an objective assessment and screening of creative potential using UUT and JST items in TTCT-Verbal.

Our first and second research questions inquired if the automated scoring of UUT and JST items across Forms A and B correlate with TTCT-manual-based scores and Snapshot originality. We found that both UUT and JST items across Forms A and B correlated significantly with both TTCT-manual-based and Snapshot originality scores. The correlations of the automated text-mining scores with the total automated test-mining scores and Snapshot originality were "strong" ($r > .50$) according to Cohen's (1988) guidelines. They were small to moderate when averaged text-mining scores from GLoVe were considered ($r = .05– .33$). This is a promising finding in several ways. First, most research on automated scoring has focused on UUT-like tasks (e.g., Beaty et al., 2014; Dumas & Dunbar, 2014; Forster & Dunbar, 2009; Forthmann, Wilken, et al., 2019; Hass, 2017a). However, using UUT alone in gifted identification may only cover a limited scope of creative potential (Runco et al., 2016). Only a few previous studies have experimented with LSA-based scoring of the Instances task (Hass, 2017b), which did not correlate with Snapshot originality of the responses. LaVoie et al. (2020) applied LSA-based scoring to the Consequences test using Reuters News corpora and found a high level of agreement with human raters. We extended this line of work to the JST items from TTCT-Verbal and found that text-mining methods do apply to them, as well.

Suitability of JST for automated scoring is important because use of a single type of divergent thinking task is not ideal (Acar & Runco, 2019; Runco et al., 2016) and tasks such as Consequences/JST tend to elicit somewhat different cognitive operations than the UUT. The cognitive load is higher with the former than the latter (Forthmann et al., 2017) and, probably because of this, latency (response time) is higher for the first response than the latter (Hass & Beaty, 2018). Silvia (2011) argued that the Alternate/Unusual Uses tasks are likely to engage executive functions, whereas the Instances (which was not investigated in the present study) are more associational. It is likely that the JST/Consequences tasks are more demanding of cause-of-effect relationships because the respondents are compelled to apply analytical thinking to hypothetical and often nonrealistic situations. The inclusion of the JST in creativity assessment is particularly important in gifted identification because hypothetical thinking is the defining feature of an intelligent mind in that it calls for "the ability to entertain an idea without accepting it" and involves imagination via the use of pretend scenarios (Amsel, 2011, p. 86). Children with high intellectual and creative potential are more likely to employ this skill as it allows thinking of alternatives to the existing reality.

From a psychometric and practical standpoint, previous research (Hass et al., 2018; Silvia, 2011; Silvia et al., 2008) reported lower interagreement reliability with Consequences than for the Alternate/Unusual Uses and this may again happen due to higher cognitive load (Forthmann et al., 2017). In the present study, JST displayed lower interagreement reliability (.71 and .69) compared with the UUT (.76 and .78). The capability of automated scoring of JST, therefore, solves multiple problems such as reliability of the scores facing creativity assessment beyond gifted identification.

When the correlations of the automated scores were compared between the two criteria (i.e., TTCT-manual-based and Snapshot originality) across UUT and JST, originality scores from both UUT and JST items correlated very highly with TTCT-manual-based scores, but the correlations with Snapshot originality were higher for UUT than JST. This may be due to the differences in the number of words in the responses generated for the UUT and JST. The responses given for the UUT were shorter than those for the JST, and longer responses are harder for the text-mining models to process accurately (Dumas et al., 2020). Still, the correlations achieved in this study (above .50) on both JST items suggest strong relationships.

In the present study, we reported the results from three automated text-mining-based scoring systems and compared aggregation method (total vs. average) against two criteria (TTCT-manual-based vs. Snapshot scoring). When the total originality scores are considered, all three scoring methods worked fairly well across the UUT and JST items. The differences among the correlations across different scoring systems favored GLoVe and TASA over EN 100k. The superiority of the GLoVE was clearest when average originality scores are considered. For the UUT items, the GLoVe-based originality provided the strongest correlation with both criteria except for one case (i.e., Averaged TASA-rated originality for UUT-Tin Cans/Form B). It also stood out as the only automated scoring method that correlated significantly with the JST items' averaged originality scores.

The superiority of the GLoVe system can be explicated by its training on a larger corpus (840 billion words) compared with TASA (37,651 words) and EN 100k (100,000 words). In addition, the nature of the texts scraped for building the corpora varies. GLoVe includes global web-based documents, whereas TASA focuses on educational texts and EN100k consists of the most frequently used words in English based on U.K.-based online platforms. Thus, GLoVe is superior in size, scope, and diversity of domains included. The inclusion of the web-based documents is certainly a big advantage as they might better approximate everyday contemporary vocabulary. Although EN 100k is also web-based, it is limited to .uk domains. Another distinctive feature of the GLoVe is that it relies on localized associations among the words by focusing on a smaller space (e.g., the same page of a book or website) rather than the entire text, increasing the precision of the semantic similarity estimates. Finally, both TASA and EN 100k were the basis for LSA, whereas GLoVe is not (see Dumas et al., 2020, for their comparison). Future studies may explore possible improvements in the scoring systems by testing different text-mining models with different corpora. For example, LaVoie et al. (2020) trained their LSA model on the Reuters News corpora (100,000 paragraphs) blended with 18,000 paragraphs of text selected from Wikipedia. They also found very high correlations ($r = .94$) with human raters of the Consequences test. Different from our study, LaVoie et al. did not examine the correlations with averaged originality scores, which may be a more effective method of controlling for the fluency confound (Forthmann et al., 2020). In this study, we found the GloVe-based originality scores converge with human ratings for both UUT items and one JST item, even when fluency is accounted for by dividing originality scores by fluency. Accounting for fluency is also the likely reason for the lower observed correlation between fluency scores and averaged originality, than between fluency and total originality scores (Hocevar, 1979). Likewise, the manual-based scores had a higher correlation with the automated scores than the Snapshot originality because both manual-based scores and total automated scores are influenced by fluency. Snapshot originality, however, is less influenced by fluency because raters focus on the entire response set rather than individual responses. Fluency is more influential on originality scores when individually scored responses are aggregated by summing (Reiter-Palmon et al., 2019). Given that higher correlations were obtained when total scores were used, using total—rather than averaged—text-mining scores could be more appropriate when the purpose is to replace traditional scoring methods of the TTCT-Verbal. This use, however, will include some level of fluency confound on originality scores. This may be acceptable up to a certain point where discriminant validity is retained because of the extended effort principle (Parnes, 1961), which suggested that the path to original responses is through generating more responses. In other words, some level of correlation between fluency and originality is well founded in the extant creativity literature. When the correlations are too high with fluency (i.e., low discriminant validity between originality and fluency), averaged originality scores should be used from GLoVe rather than other text-mining methods because it has been shown to provide stronger correlations with human judgments (Dumas et al., 2020).

Another powerful finding in the present study is the amount of information provided by only two items (UUT and JST) from TTCT-Verbal when the open scoring is applied. Administering two items from TTCT-Verbal can predict 55% to 68% of total TTCT-manual-based originality scores across all six activities of the TTCT-Verbal. This evidence demonstrates that the use of UUT and JST alone may suffice as a short form when GloVe- or TASA-based automated scores are used. Note here that the TTCT-Verbal takes 40 min to administer and another 30 min to 1 hr to be scored by a human rater, who also needs to attend a full-day training and devote additional hours of practice to prove scoring consistency. Naturally, this effort needs to be compensated and adds financial burden on the respondents or the school district. Automated scoring increases the usability of the TTCT-Verbal because it saves time, saves money, and reduces subjectivity in scoring (Smith et al., 2012). Thus, the evidence we present here shows the promise of the use of those two items of TTCT-Verbal at least as a screening tool. However, further psychometric investigation of scale reliability of the full and two-item versions of TTCT-Verbal is

necessary before using the two-item version in place of the full scale. If high reliability is obtained for the short form, using the two-item form will save time or reduce the number of testing sessions. These are all beneficial for practical purposes in the context of a multiple-criteria approach to gifted identification.

Our findings also question the need for scoring TTCT-Verbal for all three indices (fluency, flexibility, and originality) when it is possible that a single index (total originality score) seems to strongly predict the aggregate creativity index combining all three scores. Again, our originality scores predicted around 70% of the variance in the TTCT-manual-based creativity scores almost instantly when the responses were given in digital format. The problem of fluency confound, which revealed itself through very high correlations ($r = .85$) among the indices (Forthmann et al., 2020) as well as activity-specific factor loadings (Krumm et al., 2016; Said-Metwaly et al., 2021), indicates the redundancy of the daunting manual scoring process that attempts to capture all three indices. Although scoring all the tasks for all three indices makes theoretical sense and better mirrors the theoretical complexity of creativity as a construct, the current widely utilized method of scoring seems to fall short of reflecting that complexity. Given this high correlation among the indices, the use of automated scoring of originality seems to be a sufficient (and even more optimal) scoring approach than scoring for all three indices.

Limitations and Future Directions

First, in the present study, we focused on only two tasks in TTCT-Verbal, the UUT and JST, the likes of which have been tested in previous research (Dumas & Dunbar, 2014; LaVoie et al., 2020). The applicability of other items and item-types should be explored in future research. This is, however, a more challenging task because except for one (i.e., "Unusual Questions"), they involve the use of an image as the stimuli rather than a word or situation described in words. Second, the automated scoring worked for just one of the JST items when averaged originality scores were used although it worked fine with both items for the total scores. This is a problem only when fluency is also scored, which appears to be unnecessary given the high correlation between the two indices. Alternative methods of automated scoring that apply to both items should be explored in future research. Based on current evidence, we recommend using TTCT-Verbal Form B if automated scoring procedures will be used.

Third, gifted identification is typically conducted at elementary school ages, whereas our sample was college students. Although age-related differences typically have been observed in the figural divergent thinking tasks but not in the verbal ones (Said-Metwaly et al., 2021), a direct application of the existing text-mining systems to children's data may provide suboptimal efficiency in capturing children's creativity. TTCT's scoring procedures are the same for all ages, but the text-mining model results may be influenced by the corpus or text-mining model employed. Indeed, an approach that is more robust to age would be to use automated scoring of children's data using a corpus fine-tuned on children's or child-directed language. One such effort is currently underway (Organisciak et al., 2021). Future studies need to examine the applicability of the presented automated scoring methods with children's data, which may need to be assessed based on a corpus developed for children. With that said, our preliminary analyses with children's responses to a different type of divergent thinking tasks show that the difference between children's and adults' corpora is negligible (at least for Grades 3–5). Therefore, the methods (i.e., the text-mining model and corpora) used in the present study are likely to be useful for children's data, as well. Fourth, performance on TTCT-Verbal may be influenced by verbal skills of the students and this may penalize students with lower vocabulary or literacy skills unless the obtained automated scores are pooled by sociodemographic characteristics (Lee et al., 2020). Alternatively, it would be safe to use instruments such as verbal measures of cognitive ability (e.g., the verbal battery of the Cognitive Abilities Test) to control for this confounding effect of verbal skills before making high-stakes decisions. Fifth, when data collection is conducted via computer where children type in their responses instead of the traditional paper–pencil method, automated scoring requires extra attention for correcting typos. Last, the text-mining models used in the Open Scoring Platform are applicable in English and future studies should explore extending it to other languages.

Conclusion

There is a large consensus on the importance of creativity in gifted education and the construct of giftedness, yet its measurement and identification in gifted identification faces some logistical challenges such as the cost of popular assessment tools and time of administering and scoring. In the present study, we indicated that such concerns can be overcome by adopting recent developments in creativity assessment: specifically, scoring responses for originality using text-mining models that generate reliable metrics of semantic distance in two of the activities in TTCT-Verbal. We found that text-mining methods can be extended beyond UUT and can be applied to JST, too. Moreover, the use of these automated scoring methods in only two activities can be indicative of performance on total scores. This is a promising sign that creativity assessment could be conducted in a cost-effective and faster way and increase the adoption of creativity assessment in gifted identification in the near future.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Institute of Education Sciences (IES) (Grant No. R305A200519).

Open Science Disclosure Statement

The data analyzed in this study are not available for purposes of reproducing the results. The code or protocol used to generate the findings reported in the article are not available for purposes of reproducing the results or replicating the study. There were no newly created, unique materials used to conduct the research.

ORCID iD

Selcuk Acar https://orcid.org/0000-0003-4044-985X

References

Acar, S., & Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal*, *26*(2), 229–238. https://doi.org/10.1080/10400419.2014.901095

Acar, S., & Runco, M. A. (2019). Divergent thinking: New methods, recent research, and extended theory. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 153–158. https://doi.org/10.1037/aca0000231

Acar, S., Runco, M. A., & Park, H. (2020). What should people be told when they take a divergent thinking test? A meta-analytic review of explicit instructions for divergent thinking. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(1), 39–49. https://doi.org/10.1037/aca0000256

Acar, S., Sen, S., & Cayirdag, N. (2016). Consistency of the performance and nonperformance methods in gifted identification: A multilevel meta-analytic review. *Gifted Child Quarterly*, *60*, 81–101. https://doi.org/10.1177%2F0016986216634438

Amsel, E. (2011). Hypothetical thinking in adolescence: Its nature, development, and applications. In J. Smetana & E. Amsel (Eds.), *Adolescence: Vulnerabilities and opportunities* (pp. 86–113). Cambridge University Press. https://doi.org/10.1017/CB09781139042819.007

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*(4), 309–319. https://doi.org/10.1037/a0029171

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 & Cognition*, *42*(7), 1186–1197. https://doi.org/10.3758/s13421-014-0428-8

Beketayev, K., & Runco, M. A. (2016). Scoring divergent thinking tests by computer with a semantics-based algorithm. *Europe's Journal of Psychology*, *12*(2), 210–220. https://doi.org/10.5964/ejop.v12i2.1127

Castejón, J. L., Gilar, R., Miñano, P., & González, M. (2016). Latent class cluster analysis in exploring different profiles of gifted and talented students. *Learning and Individual Differences*, *50*, 166–174. https://doi.org/10.1016/j.lindif.2016.08.003

Chen, P., Lin, S., & Chu, Y. (2011). Using Google latent semantic distance to extract the most relevant information. *Expert Systems With Applications*, *38*, 7349–7358. https://doi.org/10.1016/j.eswa.2010.12.092

Christensen, P. R., Merrifield, P. R., & Guilford, J. P. (1958). *Consequences: Manual for administration, scoring, and interpretation*. Sheridan Supply.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Lawrence Erlbaum.

Diedrich, J., Benedek, M., Jauk, E., & Neubauer, A. C. (2015). Are creative ideas novel and useful? *Psychology of Aesthetics, Creativity, and the Arts*, *9*(1), 35–40. https://doi.org/10.1037/a0038688

Dumas, D. (2018). Relational reasoning and divergent thinking: An examination of the threshold hypothesis with quantile regression. *Contemporary Educational Psychology*, *53*, 1–14. https://doi.org/10.1016/j.cedpsych.2018.01.003

Dumas, D., & Dunbar, K. N. (2014). Understanding fluency and originality: A latent variable perspective. *Thinking Skills and Creativity*, *14*, 56–67. https://doi.org/10.1016/j.tsc.2014.09.003

Dumas, D., & Dunbar, K. N. (2016). The creative stereotype effect. *PLOS ONE*, *11*(2), Article e0142567. https://doi.org/10.1371/journal.pone.0142567

Dumas, D., Organisciak, P., & Doherty, M. (2020). Measuring divergent thinking originality with human raters and text-mining models: A psychometric comparison of methods. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. https://doi.org/10.1037/aca0000319

Dumas, D., & Runco, M. (2018). Objectively scoring divergent thinking tests for originality: A re-analysis and extension. *Creativity Research Journal*, *30*(4), 466–468. https://doi.org/10.1080/10400419.2018.1544601

Dumas, D. G., & Strickland, A. L. (2018). From book to bludgeon: A closer look at unsolicited malevolent responses on the alternate uses task. *Creativity Research Journal*, *30*(4), 439–450. https://doi.org/10.1080/10400419.2018.1535790

Forthmann, B., Holling, H., Zandi, N., Gerwig, A., Çelik, P., Storme, M., & Lubart, T. (2017). Missing creativity: The effect of cognitive workload on rater (dis-) agreement in subjective divergent-thinking scores. *Thinking Skills and Creativity*, *23*, 129–139. https://doi.org/10.1016/j.tsc.2016.12.005

Forthmann, B., Oyebade, O., Ojo, A., Günther, F., & Holling, H. (2019). Application of latent semantic analysis to divergent thinking is biased by elaboration. *The Journal of Creative Behavior*, *53*(4), 559–575. https://doi-org.libproxy.library.unt.edu/10.1002/jocb.240

Forthmann, B., Szardenings, C., & Holling, H. (2020). Understanding the confounding effect of fluency in divergent thinking scores: Revisiting average scores to quantify artifactual correlation. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(1), 94–112. https://doi.org/10.1037/aca0000196

Forthmann, B., Wilken, A., Doebler, P., & Holling, H. (2019). Strategy induction enhances creativity in figural divergent thinking. *The Journal of Creative Behavior*, *53*(1), 18–29. https://doi.org/10.1002/jocb.159

Forster, E. A., & Dunbar, K. N. (2009). Creative evaluation through latent semantic analysis. *Proceedings of the Annual Conference of the Cognitive Science Society*, *2009*, 602–607.

Fox, C. (1989, September). A stop list for general text. *ACM SIGIR Forum*, *24*(1–2), 19–21. https://doi.org/10.1145/378881.378888

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological*

Forecasting & Social Change, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Gagné, F. (2005). *From gifts to talents: The DMGT as a developmental model*. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 98–120). Cambridge University Press.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*(2), 117–135. https://doi.org/10.1016/j.jsp.2006.05.005

Gray, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S., Patrick, J., Huang, L., Kenett, Y. N., & Lewis, K. (2019). "Forward flow": A new measure to quantify free thought and predict creativity. *The American Psychologist*, *74*, 539–554. http://dx.doi.org/10.1037/amp0000391

Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*, *2*(1), 1–25. https://doi.org/10.1177/2332858415622175

Grissom, J. A., Redding, C., & Bleiberg, J. F. (2019). Money over merit? Socioeconomic gaps in receipt of gifted services. *Harvard Education Review*, *89*(3), 337–369. https://doi.org/10.17763/1943-5045-89.3.337

Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun-An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*(4), 930–944. https://doi.org/10.3758/s13428-014-0529-0

Hass, R. W. (2015). Feasibility of online divergent thinking assessment. *Computers in Human Behavior*, *46*, 85–93. https://doi.org/10.1016/j.chb.2014.12.056

Hass, R. W. (2017a). Semantic search during divergent thinking. *Cognition*, *166*, 344–357. https://doi.org/10.1016/j.cognition.2017.05.039

Hass, R. W. (2017b). Tracking the dynamics of divergent thinking via semantic distance: Analytic methods and theoretical implications. *Memory & Cognition*, *45*(2), 233–244. https://doi.org/10.3758/s13421-016-0659-y

Hass, R. W., & Beaty, R. E. (2018). Use or consequences: Probing the cognitive difference between two measures of divergent thinking. *Frontiers in Psychology*, *9*, 2327. https://doi.org/10.3389/fpsyg.2018.02327

Hass, R. W., Rivera, M., & Silvia, P. J. (2018). On the dependability and feasibility of layperson ratings of divergent thinking. *Frontiers in Psychology*, *9*, Article 1343. https://doi.org/10.3389/fpsyg.2018.01343

Heinen, D. J., & Johnson, D. R. (2018). Semantic distance: An automated measure of creativity that is novel and appropriate. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(2), 144–156. https://doi.org/10.1037/aca0000125

Hocevar, D. (1979). Ideational fluency as a confounding factor in the measurement of originality. *Journal of Educational Psychology*, *71*(2), 191–196. https://doi.org/10.1037/0022-0663.71.2.191

Hunsaker, S. L., & Callahan, C. M. (1995). Creativity and giftedness: Published instrument uses and abuses. *Gifted Child Quarterly*, *39*(2), 110–114. https://doi.org/10.1177/001698629503900207

Jauk, E., Benedek, M., Dunst, B., & Neubauer, A. C. (2013). The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence*, *41*(4), 212–221. https://doi.org/10.1016/j.intell.2013.03.003

Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D. M., Gajda, A., Chruszczewski, M. H., & Benedek, M. (2016). Is creativity without intelligence possible? A necessary condition analysis. *Intelligence*, *57*, 105–117. https://doi.org/10.1016/j.intell.2016.04.006

Kaufman, J. C., Plucker, J. A., & Russell, C. M. (2012). Identifying and assessing creativity as a component of giftedness. *Journal of Psychoeducational Assessment*, *30*(1), 60–73. https://doi.org/10.1177/0734282911428196

Kim, K. H., Kaufman, J. C., Baer, J., & Sriraman, B. (Eds.). (2013). *Creatively gifted students are not like other gifted students: Research, theory, and practice* (Vol. 5). Springer Science & Business Media.

Krathwohl, D. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, *41*, 212–218. https://doi.org/10.1207/s15430421tip4104_2

Krisel, S. C., & Cowan, R. S. (1997). Georgia's journey toward multiple-criteria identification of gifted students. *Roeper Review*, *20*, A1–A3. https://doi.org/10.1080/02783199709553867

Krumm, G., Aranguren, M., Arán Filippetti, V., & Lemos, V. (2016). Factor structure of the Torrance Tests of Creative Thinking Verbal Form B in a Spanish-speaking population. *Journal of Creative Behavior*, *50*, 150–164. http://dx.doi.org/10.1002/jocb.76

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

LaVoie, N., Parker, J., Legree, P. J., Ardison, S., & Kilcullen, R. N. (2020). Using latent semantic analysis to score short answer constructed responses: Automated scoring of the consequences test. *Educational and Psychological Measurement*, *80*(2), 399–414. https://doi.org/10.1177/0013164419860575

Lee, L. E., Ottwein, J. K., & Peters, S. J. (2020). *Eight universal truths of identifying students for advanced academic interventions*. In J. L. Jolly & J. H. Robins (Eds.), *Methods & materials for teaching the gifted* (5th ed., pp. 61–79). Prufrock Press.

Lichtenberg, J., Woock, C., & Wright, M. (2008). *Ready to innovate: Are educators and executives aligned on the creative readiness of the U.S. workforce?* (Conference Board, Research Report 1424). Conference Board.

Luria, S. R., O'Brien, R. L., & Kaufman, J. C. (2016). Creativity in gifted identification: Increasing accuracy and diversity. *Annals of the New York Academy of Sciences*, *1377*(1), 44–52. https://doi.org/10.1111/nyas.13136

Marland, S. P., Jr. (1972). *Education of the gifted and talented: Report to the Congress of the United States by the U.S* (Government Documents, Y4.L 11/2: G36). U.S. Government Printing Office.

Matthews, M. S. (2015). Creativity and leadership's role in gifted identification and programming in the USA: A pilot study. *Asia Pacific Education Review*, *16*(2), 247–256. https://doi.org/10.1007/s12564-015-9373-x

McBee, M. T., Peters, S. J., & Miller, E. M. (2016). The impact of the nomination stage on gifted program identification: A comprehensive psychometric analysis. *Gifted Child Quarterly*, *60*, 258–278. https://doi.org/10.1177/0016986216656256

Mun, R. U., Hemmler, V., Langley, S. D., Ware, S., Gubbins, E. J., Callahan, C. M., McCoach, B., & Siegle, D. (2020). Identifying and serving English Learners in gifted education: Looking back and moving forward. *Journal for the Education of the Gifted*, *43*(4), 297–335. https://doi.org/10.1177/0162353220955230

Organisciak, P. & Dumas, D. (2020). *Open creativity scoring* [Computer software]. University of Denver.

Organisciak, P., Ryan, M., Newman, M., Acar, S., & Dumas, D. (2021). *MOTES corpus: A text-mining corpus for children's language*. [Manuscript in preparation].

Parnes, S. J. (1961). Effects of extended effort in creative problem solving. *Journal of Educational Psychology*, *52*(3), 117–122. https://doi.org/10.1037/h0044650

Partnership for 21st Century Skills. (2006). *A state leader's action guide to 21st century skills: A new vision for education*.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In A. Moscitti, A. Pang, & B. Daelemans (Eds.), *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1532–1543). Association for Computational Linguistics.

Peters, S. J., Gentry, M., Whiting, G. W., & McBee, M. T. (2019). Who gets served in gifted education? Demographic proportionality and a call for action. *Gifted Child Quarterly*, *63*(4), 273–287. https://doi.org/10.1177/0016986219833738

Peters, S. J., Ottwein, J. K., Lee, L. E., & Matthews, M. S. (2020). *Identification*. In J. A. Plucker & C. Callahan (Eds.), *Critical issues & practices in gifted education* (3rd ed., pp. 261–272). Prufrock Press.

Petrone, P. (2019). *Why creativity is the most important skill in the world*. https://learning.linkedin.com/blog/top-skills/why-creativity-is-the-most-important-skill-in-the-world

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. http://dx.doi.org/10.3758/s13428-013-0401-7

Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 144–152. https://doi.org/10.1037/aca0000227

Renzulli, J., & Reis, S. (2014). *The schoolwide enrichment model: A how-to guide for talent development*. Sourcebooks.

Renzulli, J. S. (2005). *The three-ring definition of giftedness: A developmental model for promoting creative productivity*. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 246–280). Cambridge University Press.

Rinn, A. N., Mun, R. U., & Hodges, J. (2020). *2018-2019 State of the states in gifted education*. National Association for Gifted Children and the Council of State Directors of Programs for the Gifted. https://www.nagc.org/2018-2019-state-states-gifted-education

Runco, M. A. (1986). Flexibility and originality in children's divergent thinking. *The Journal of Psychology*, *120*(4), 345–352. https://doi.org/10.1080/00223980.1986.9712632

Runco, M. A., Abdulla, A. M., Paek, S. H., Al-Jasim, F. A., & Alsuwaidi, H. N. (2016). Which test of divergent thinking is best? *Creativity. Theories–Research-Applications*, *3*(1), 4–18. https://doi.org/10.1515/ctra-2016-0001

Runco, M. A., & Mraz, W. (1992). Scoring divergent thinking tests using total ideational output and a creativity index. *Educational and Psychological Measurement*, *52*(1), 213–221. https://doi.org/10.1177/001316449205200126

Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., & Barbot, B. (2021). Does the fourth-grade slump in creativity actually exist? A Meta-analysis of the development of divergent thinking in school-age children and adolescents. *Educational Psychology Review*, *33*, 275–298. https://doi.org/10.1007/s10648-020-09547-9

Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., & Van den Noortgate, W. (2020). Testing conditions and creative performance: Meta-analyses of the impact of time limits and instructions. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(1), 15–38. https://doi.org/10.1037/aca0000244

Silverman, L. K., & Gilman, B. J. (2020). Best practices in gifted identification and assessment: Lessons from the WISC-V. *Psychology in the Schools*, *57*(10), 1569–1581. https://doi.org/10.1002/pits.22361

Silvia, P. J. (2011). Subjective scoring of divergent thinking: Examining the reliability of unusual uses, instances, and consequences tasks. *Thinking Skills and Creativity*, *6*(1), 24–30. https://doi.org/10.1016/j.tsc.2010.06.001

Silvia, P. J., Martin, C., & Nusbaum, E. C. (2009). A snapshot of creativity: Evaluating a quick and simple method for assessing divergent thinking. *Thinking Skills and Creativity*, *4*(2), 79–85. https://doi.org/10.1016/j.tsc.2009.06.005\

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*(2), 68–85. https://doi.org/10.1037/1931-3896.2.2.68

Smith, G. T., Combs, J. L., & Pearson, C. M. (2012). *Brief instruments and short forms* In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbooks in psychology®. APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (pp. 395–409). https://doi.org/10.1037/13619-021

Sternberg, R. J. (2003). *WICS: Wisdom, intelligence, and creativity, synthesized*. Cambridge University Press.

Subotnik, R. F., & Jarvin, L. (2005). Beyond expertise: Conceptions of giftedness as great performance. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 343–357). http://dx.doi.org/10.1017/CBO9780511610455.020

Torrance, E. P. (2008). *The Torrance tests of creative thinking norms—Technical manual figural (streamlined) forms A & B*. Scholastic Testing Service.

Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020). A reappraisal of the threshold hypothesis of creativity and intelligence. *Journal of Intelligence*, *8*(4), Article 38. https://doi.org/10.3390/jintelligence8040038

Wild, K., Kilgarriff, A., & Tugwell, D. (2013). The Oxford Children's Corpus: Using a children's corpus in lexicography. *International Journal of Lexicography*, *26*(2), 190–218. https://doi.org/10.1093/ijl/ecs017

Wiley, J., & Voss, J. F. (1996). The effects of "playing historian" on learning in history. *Applied Cognitive Psychology*, *10*(7), 63–72. https://doi.org/10.1002/(SICI)1099-0720(199611)10:7<63::AID-ACP438>3.0.CO;2-5

Zenasni, F., Mourgues, C., Nelson, J., Muter, C., & Myszkowski, N. (2016). How does creative giftedness differ from academic giftedness? A multidimensional conception. *Learning and Individual Differences*, *52*, 216–223. http://dx.doi.org/10.1016/j.lindif.2016.09.003

Author Biographies

Selcuk Acar is an associate professor of educational psychology at the University of North Texas. He received his PhD in educational psychology (with an emphasis in gifted and creative education) from the University of Georgia. His primary area of research interest includes divergent thinking, assessment of creativity, and the education of the gifted and talented. He has worked or studied at two different creativity centers: Torrance Center for Creativity and Talent Development of University of Georgia, and the International Center for Studies in Creativity (ICSC) of SUNY Buffalo State. He is currently leading a 3-year project to develop a new measure of original thinking (MOTES) funded by the Institute of Education Sciences. Besides his role as the associate editor of *Journal of Creativity*, he is serving on the editorial/review board of *Journal of Creative Behavior, Creativity Research Journal*, and the *Journal of Advanced Academics*.

Kelly Berthiaume is a postdoctoral research associate at the University of North Texas. She received her PhD in human development and family science from Florida State University. Her primary area of research interest focuses on how children's development of beliefs about the nature of learning-related cognitions and creativity are influenced by parents' socialization practices and contextual factors such as community type. As a certified family life educator, she approaches research through the lens of translational science by utilizing evidence and theory as a foundation to inform the design, implementation, and evaluation of an effective evidence-based parent education program that educates parents about influential parenting practices and behavior that positively impact children's motivational framework and cognitive development as well as the parent–child relationship quality.

Katalin Grajzel is a PhD student at the University of Denver completing her degree in research methods and statistics with focus on psychometrics. She holds a master's degree in psychology and in counseling. She has worked for the U.S. Government supporting research on psychological well-being and recovery of military personnel and worked as a counselor with underprivileged populations. Her primary research interests include the assessment of creativity using divergent thinking tasks and employing text-mining approaches for quantification answers on these tasks. She is a research assistant for the MOTES project focusing on measurement-related tasks such as item development, reliability, validity, modeling, and so on.

Denis Dumas is an assistant professor of research methods and statistics at the University of Denver's Morgridge College of Education. In general, his work focuses on understanding student learning, cognition, and creativity through the application and refinement of latent variable methods, especially multidimensional item response theory and nonlinear growth models. He completed his doctoral work in educational psychology, and master's degree in educational measurement and statistics, at the University of Maryland–College Park, and was a postdoctoral researcher at the American Educational Research Association. His work has also been previously funded by the National Academy of Education, Spencer Foundation, Institute of Educational Sciences, and the Hewlett Foundation. He was among the first researchers to apply text-mining-based methodologies to the measurement of creative thinking and is excited to refine and improve these processes in the context of elementary education as a member of the MOTES project.

Charles "Tedd" Flemister is a PhD student at the University of North Texas. He is currently pursuing his PhD in educational psychology with a concentration in gifted and talented education. Previously, he received his master's in education and teaching certificate in secondary social studies at the University of North Texas. For the MOTES project, he is a research assistant, who works primarily with assessment design and illustration as well as data collection.

Peter Organisciak is an assistant professor of library and information science at the University of Denver, with a focus on computation text analysis in the context of human factors and digital library research. Beyond MOTES, his recent work deals with machine learning methods for relationship extraction in archival collections. He holds a PhD in library and information science (UIUC) and previously worked on scholarly access to millions of digitized books at the HathiTrust Research Center. His work has been funded by the Institute for Library and Museum Services, National Endowment for the Humanities, and the Institute of Education Sciences and has received paper awards from the Association for the Advancement of Artificial Intelligence (AAAI) and the Association for Information Science and Technology (ASIS&T), and the iSchools Consortium.

Manuscript received: February 28, 2021; Final revision received: October 14, 2021; Accepted: November 3, 2021.