JOURNAL OF LEARNING ANALYTICS

Volume 12(2), 67–82. https://doi.org/10.18608/jla.2025.8857

Disentangling the Influences of Creativity and Verbal Fluency on Writing

Lauren E. Flynn[1] and Laura K. Allen[2]

Abstract

This study investigates how individual differences related to creative and verbal fluencies can be differentiated and modelled through the linguistic features that manifest in student writing. Using computational linguistics, we analyzed samples of intrapersonal writing to extract linguistic features that align to specific dimensions of language. Creative fluency scores were calculated based on the number of unique ideas generated during the Alternate Uses Task (AUT); verbal fluency scores were calculated based on the number of unique semantic categories produced during the animal naming task. We then developed machine learning models to predict creative and verbal fluency scores based on the linguistic features of participant essays. Results indicate that creative fluency is more strongly predicted by linguistic features (particularly descriptive and cohesion indices) than verbal fluency. Key predictors of creative fluency include features that capture lexical diversity and the global connectedness of ideas. Importantly, our results align with theoretical frameworks related to convergent and divergent thinking. Findings highlight the potential for leveraging learning analytics to offer new insights into complex cognitive processes such as creativity. Implications for stealth assessments and personalized feedback within automated writing systems are discussed as paths for future research.

Notes for Practice

- Creativity is a vital 21st-century skill for students to learn in preparation to successfully entering the workforce.
- Cognitive researchers have traditionally conflated creativity and verbal fluency. This study aims to differentiate these constructs in the context of writing.
- Writing serves as a popular expression of creativity but relatively little is known about how creativity manifests itself within the final written product, nor with intrapersonal communicative writing.
- The current study explores how individual differences in two distinct types of fluency (creative fluency, verbal fluency) manifest differently within the linguistic dimensions of writing using machine learning and computational linguistics, utilizing natural language processing (NLP) indices measured from the Writing Assessment Tool (WAT).
- Key findings: Our study reveals that the linguistic properties of writing can be modelled more accurately by one's creativity than verbal fluency — indicating distinct differences in how these skills manifest within written intrapersonal text.
- Practical applications: Our results have meaningful implications for the integration of individualized feedback in automated writing evaluation (AWE) and intelligent tutoring systems (ITS), where student writing could serve as a form of stealth assessment to infer individual differences in creative and verbal fluency skills.

Keywords: Writing, creativity, natural language processing, verbal fluency

Submitted: 05/01/2025 — **Accepted:** 14/07/2025 — **Published:** 29/08/2025

Corresponding author [1]Email: flynn598@umn.edu Address: Department of Educational Psychology, University of Minnesota, 250 Education Sciences Building, 56 East River Road, Minneapolis, MN 55455, USA. ORCID iD: https://orcid.org/0000-0001-9359-6322
[2]Email: allenl@umn.edu Address: Department of Educational Psychology, University of Minnesota, 250 Education Sciences Building, 56 East River Road, Minneapolis, MN 55455, USA. ORCID iD: https://orcid.org/0000-0001-5582-1402

1. Introduction

Creativity, referred to by some as the ability to create original and appropriate ideas (Runco & Jaeger, 2012), is now widely recognized as an important 21st-century skill for developing critical skills such as metacognition, open-mindedness, collaborative problem-solving, and mental flexibility (Grey & Morris, 2024; OECD, 2024). Thus, the modern education system that serves to instill these 21st-century skills in future generations of students must promote creativity so that students can learn how to quickly adapt within a constantly changing world (OECD, 2024). However, research examining student creativity currently lacks external validity and can require tedious coding of creativity assessments. The minimal work addressing how students can strengthen their creative abilities is partially due to the many challenges in conceptualizing the multidimensional nature of creativity and measuring it in a valid and reliable manner (Davis, 2004; Rogaten & Moneta, 2016; Runco & Pritzker, 2011). It is thus crucial that more research examining student creativity and its influence on natural learning processes and products is essential to understanding the best practices when promoting and assessing creativity within education.

One common conceptualization of creativity is its classification between Big-C and little-c theoretical perspectives. *Big-C creativity* conceptualizes creativity as a talent or attribute of a few, special, extraordinary individuals (Davis, 2004; Rogaten & Moneta, 2016; Runco, 1995; Treffinger, 1986). This view of creativity is extremely limited though in terms of education, as most students aren't considered gifted in any specific domain, let alone old enough to be recognized as a major contributor to their field. As opposed to the natural creative abilities innate to a few individuals in Big-C creativity, little-c creativity postulates that everyone has creative potential that they can further develop with the right intervention. This view of creativity emphasizes the importance of everyday creativity to solve minor problems encountered in daily life through the use of general problem-solving skills (Richards et al., 1988; Rogaten & Moneta, 2016; Runco, 1995). Thus, promoting creativity within general education reflects its alignment with the *little-c creativity* perspective (Rogaten & Moneta, 2016).

One interesting application of one's creativity is through writing tasks that offer students the unique opportunity to apply their creative abilities. Writing tasks are already well-established within the context of education and can be used as a more naturalistic assessment of student little-c creativity (Rogaten & Moneta, 2016). Writing tasks typically involve transforming one's thoughts and ideas into coherent sentences and the final written product produced can be conceived of as little-c creativity in response to a particular problem identified in the prompt or task instructions (Carey & Flower, 1989). Additionally, writing is an important skill crucial to learning, academic achievement, reflection, and communication that is greatly impacted by one's individual differences (Sarikaya, 2024). For example, writing skills related to brainstorming, organization, and style are thought to be influenced by one's creativity. Specifically, creativity is thought to influence the writing process in terms of fluency (communicating more coherent ideas; Fredericks, 2005), flexibility (establishing connections between ideas; Pei et al., 2017), originality (generating novel content; Wati, 2019), and elaboration (building on ideas; Fredericks, 2005; Sarikaya, 2024). Thus, creativity is key in generating content, drawing connections from prior knowledge, increasing novelty, and creating a sense of flow or cohesiveness.

Despite the established link between creativity and writing, there is limited research on the effects of creativity on writing (i.e., looking at how creativity manifests in the linguistic dimensions of one's written product). Creativity is a complex, multifaceted construct that deserves a more multidimensional approach to fully capture its specific nuances. Another major gap in the current literature is that the effects between creative fluency and the more general construct of verbal fluency (i.e., focusing solely on quantifying the number of ideas one can generate) have yet to be distinguished. Whereas *creative fluency* involves the generation of unique ideas that are coherently combined, *verbal fluency* involves the production of any ideas, whether unique or not (Fredericks, 2005; Sarikaya, 2024).

To address these gaps in the field, the current study explores how creative and verbal fluencies (as assessed using traditional cognitive psychology tasks) differentially relate to real-world expressions of creativity, through writing. Specifically, we modelled scores of individual creative fluency and verbal fluency using multiple linguistic dimensions used in prior work to assess writing characteristics (i.e., descriptive, cohesion, lexical, lexical diversity, sentiment, and syntax). Linguistic properties from two intrapersonal forms of writing (i.e., spontaneous thought stream task, diary-entry task) were calculated using computational linguistics. Machine learning techniques were further leveraged to explore the complex, multidimensional relationship between language and fluency scores.

We first review the literature related to domain-general assessments of creativity, typically consisting of traditional cognitive psychology tasks that measure either convergent or divergent thinking skills, and verbal fluency. We then discuss research exploring writing as a creative product that can be assessed to predict differences in cognitive processes (i.e., creative or verbal fluency). Background information related to our computational approach using learning analytics is additionally discussed before laying out our specific research questions.

2. Background

2.1. Measuring Individual Differences Using Traditional Cognitive Approaches

2.1.1. Measuring Creativity

The process approach to creativity involves the combination of both convergent and divergent thinking skills (Gerver et al., 2023). While convergent thinking involves problem-solving with only one correct solution (Runco et al., 2010), divergent thinking involves solving complex, open-ended problems that have potentially infinite possible solutions (Gerver et al., 2023; Guilford, 1950). In other words, convergent thinking is a focused, narrow search process for one specific answer to a particular problem and divergent thinking is a broad, flexible search that encourages multiple solutions to a problem (Said-Metwaly et al., 2017). Researchers typically prioritize one of these thinking skills over the other when measuring creative ability. Based on the different definitions or conceptualizations of creativity (i.e., convergent, divergent) adopted, different cognitive tasks can be implemented (Said-Metwaly et al., 2017). Thus, the creativity task that researchers choose is heavily dependent on the definition of creativity they select (Batey, 2012; Said-Metwaly et al., 2017).

The most common task used to measure convergent creativity is the Remote Associates Test (RAT). During this task participants must think of a fourth word that conceptually links three seemingly unrelated words. For example, given the words *cake*, *cottage*, and *Swiss*, subjects need to converge on the single concept that unites them — *cheese* (Gerver et al., 2023; Mednick, 1962). Convergent thinking tasks, such as the RAT, can be solved using strategic analysis or spontaneous insight. Strategic analysis employs deliberate search strategies to find possible solutions within a given problem space (Ericsson & Simon, 1998; Gerver et al., 2023; Kounios et al., 1987; Newell & Simon, 1972). However, insight is also commonly reported during the RAT, in which a novel solution suddenly emerges into one's consciousness, resulting in what is known as an "aha" moment in the spontaneous thought literature (Gerver et al., 2023; Metcalfe & Wiebe, 1987; Ovington et al., 2018; Smith & Kounios, 1996). Typically, creative fluency is conceptualized under this task as the number of total correct solutions given.

Divergent thinking is typically assessed using the Alternate Uses Task (AUT). This task requires participants to come up with as many unique or creative uses as possible for a common household object (Gerver et al., 2023; Guilford, 1950; Guilford et al., 1960; Said-Metwaly et al., 2017; Torrance, 1972). This task has been used to assess divergent thinking and general creative ability for decades and continues to be utilized in current creativity research designs (Dumas et al., 2021; Dumas & Strickland, 2018; Guilford, 1967; Hudson, 2019; Puryear et al., 2017; Torrance, 1972). For example, given the object *mug*, participants name alternate or creative uses for it such as *using it as a plant pot*, *cutting circles into pastry dough*, or *storing pencils*. One way to assess divergent thinking tasks, such as the AUT, is to count the number of responses produced, known as *creative fluency* (Gerver et al., 2023; Runco et al., 2011). Thus, in the example given two sentences above, the fluency score would be three. This tally of creative uses is the primary way of operationalizing creative fluency in existing literature (Dumas et al., 2021; Plucker & Makel, 2010). Importantly, this task continues to be used because of its predictive validity and correlations to real-world creative achievement (Beaty et al., 2018; Gerver et al., 2023; Jauk et al., 2014). However, the psychometric underpinnings of participant AUT fluency scores is relatively underdeveloped compared to other psychometric tasks, such as personality assessments (Dumas et al., 2021; Marek & Ben-Porath, 2017), and is typically confounded with individual differences in verbal fluency (Gerver et al., 2023).

Recent advances in large language models (LLMs) have significantly enhanced the automated scoring of divergent thinking tasks. For example, Organisciak et al. (2023) developed a fine-tuned deep neural network-based LLM trained on thousands of human-judged AUT responses. This approach yielded strong correlations with human ratings and outperformed earlier automated methods that relied solely on the semantic similarity between test items and responses. Other advanced supervised techniques have also demonstrated promise, including machine learning models based on engineered linguistic features (Buczak et al., 2023) and hybrid approaches that combine clustering of semantic embeddings with human-rated scores (Organisciak et al., 2023; Stevenson et al., 2020). Overall, these developments represent a substantial improvement over traditional semantic modelling techniques, such as latent semantic analysis (LSA) and Latent Dirichlet Allocation (LDA) and offer more accurate and flexible approaches for predicting creativity from written responses.

2.1.2. Measuring Verbal Fluency

In contrast to creative fluency, in which the creative quality of responses is assessed, verbal fluency is another broader type of fluency that looks solely at the number of responses produced for a single cue, i.e., naming items that belong to a certain category. One common task used to measure verbal fluency is the animal naming task. In this task, participants are asked to list as many words within the semantic category animals as possible (Gerver et al., 2023; Shao et al., 2014). Interestingly, the effects of verbal fluency can occur either as a lower-level automatic process such as semantic resonance or higher-level strategic process, i.e., thinking of a word's synonym to use instead (Craik et al., 1996). Thus, verbal fluency is influenced by both controlled and associative cognitive processes, similar to the role of creativity (Dygert & Jarosz, 2020).

Importantly, while creative fluency relies on the number of unique responses for a cue, such as listing unique uses for an ordinary item, verbal fluency scores do not measure the uniqueness of responses. In other words, whereas creative fluency tasks emphasize unique responses that are more outside-the-box, verbal fluency tasks count the number of responses that fit in-the-box (the "box" refers here to some broad semantic category, such as animals; Gerver et al., 2023; Shao et al., 2014). Thus, while there are clear parallels between the ways that cognitive researchers measure creative and verbal fluency such as the reliance on one's semantic memory, both tasks are used to measure very different cognitive processes (i.e., automatic semantic associations measured in verbal fluency tasks vs. more strategic novel associations elicited in creativity tasks; Collins & Loftus, 1975; Gerver et al., 2023).

Although researchers can reliably measure creative and verbal fluency using these traditional cognitive tasks, relatively little is known about how these scores manifest within real-world creative contexts (Gerver et al., 2023). One context that lends itself particularly well to the study of creativity, and can be studied using computational learning analytics, is writing (Carey & Flower, 1989). Examining writing is a naturalistic method of assessing student creativity in the classroom and acts as both a practical skill necessary for education and a final creative product. There is also relatively limited research on differentiating the effects between creativity and verbal fluency, a distinction that may be crucial to the identification of students with creative potential and thus the amount of and/or types of interventions offered to them. Therefore, the current study leverages a computational approach to explore how individual differences in creative and verbal fluency manifest differently within various linguistic dimensions of writing, specifically *intrapersonal writing*. Using advanced computational linguistics on student writing allows us to uncover nuances between the linguistic signatures of creativity and verbal fluency within writing. Thus, our methodological approach offers a novel way to examine the relationships between creativity, verbal fluency, and writing characteristics.

2.2. Leveraging Learning Analytics to Capture the Multidimensionality of Writing

2.2.1. Measuring Creative Expression Through Intrapersonal Writing

The writing process is multidimensional — involving many interwoven cognitive processes such as idea generation, lexical retrieval, and working memory (McCutchen et al., 1994; Öncel et al., 2021). Other important variables that affect writing success include individual differences such as vocabulary knowledge, self-regulation skills, and one's motivation (Sarikaya, 2024). Among such influential factors is creativity and its relation to idea generation, organization, and writing style (Sarikaya, 2024). Writers use these processes as they move dynamically between phases of content generation and rhetorical development (Bereiter & Scardamalia, 1987).

The informal nature and lack of constraints involved within *intrapersonal writing*, in particular, makes it a popular outlet for creative expression (Bower et al., 2023; Carey & Flower, 1989). This stylistic form of writing involves self-directed communication, in which the sender and receiver of information are contained to a single individual (C. Bainbridge, personal communication; Ruesch & Bateson, 2008). This type of writing is also considered to be very expressive, eliciting states of reflection, and aligning with how experts in education are currently assessing student creativity through products of written creative expression (OECD, 2024). Writing tasks, in general, also serve as "ill-defined problems" that require students to develop creative "solutions" they must then communicate through writing (Carey & Flower, 1989; Simon, 1985).

Writing entries produced in the current study serve as the physical manifestation of participants' individual skills related to creative and verbal fluency. Two distinct types of intrapersonal writing tasks were elicited — one in which students simply relay the spontaneous thoughts that flow through their consciousness, and the other that stimulates reflection through diary writing. Collecting these types of writing samples provides rich insight into how participants generate, connect, and organize the ideas they express across multiple dimensions of language.

2.2.2. The Multiple Dimensions of Language

Viewing writing as an ill-defined problem-solving task, writers must construct and elaborate on their mental representation of the task, integrate their relevant topic and rhetorical prior knowledge, and activate creative problem-solving skills to translate their thoughts and ideas into words on a page (Carey & Flower, 1989). Thus, creativity appears to be an important asset that enhances the ease with which students can generate new ideas and act on those ideas flexibly (McCutchen et al., 1994). However, work examining the effects of creativity on the final written product are extremely limited, and have traditionally only compared assessments of creativity with overall writing quality (McCutchen et al., 1994) or the linguistic features present within the specific task of creative writing (Zedelius et al., 2019). Additionally, skilled writers tend to be more verbally fluent and faster at lexical retrieval during the translation process, aiding in online sentence generation (McCutchen et al., 1994). Thus, the current study examines how creativity and verbal fluency manifest and distinguish themselves in writing by examining the linguistic features of the written product.

Recent advances in natural language processing (NLP) allow for more complex, multidimensional analyses of writing, the creative product produced in the current study. These techniques automatically calculate information about the linguistic and stylistic properties of the text that can be used to glean insights into the individual differences of the writer (Crossley, Allen, Snow & McNamara, 2016; Fang et al., 2023). Thus, in the current study, writing tasks are used to elicit products of creative expression from students. Computational linguistics can then be leveraged to compare different linguistic dimensions to scores on traditional measures of creativity and verbal fluency. Importantly, capturing the multidimensionality of language using NLP is the first necessary step in differentiating the effects of complex individual differences, such as creative and verbal fluency, from subtle linguistic markers embedded in one's language use. Using automated linguistic analysis of student writing has proven to be an effective way to capture these subtle linguistic differences (Allen & McNamara, 2015; Crossley, Allen Snow & McNamara, 2016; Fang et al., 2023; Öncel et al., 2021). For example, Allen and McNamara (2015) found that the lexical properties of students' independent essays accounted for 44% of the variance in vocabulary knowledge. Findings such as this suggest that one's language serves as a proxy for individual differences (Crossley et al., 2014; Crossley, Allen Snow & McNamara, 2016).

The current study utilizes the Writing Assessment Tool (WAT) to capture multiple dimensions of writing (i.e., descriptive, lexical, lexical diversity, cohesion, syntactic, sentiment) that may vary in relation to students' verbal and/or creative fluency scores. The development of WAT has allowed researchers to calculate hundreds of linguistic indices automatically for batches of writing samples, such as corpora of text documents. WAT uses NLP to analyze each text across a series of linguistic dimensions (Crossley, Allen, Snow, & McNamara, 2016). Importantly, this tool was created using a computational linguistics approach in which the indices measured are each theoretically tied to the constructs that characterize each linguistic dimension (McNamara et al., 2019; Öncel et al., 2021). Thus, this tool has been utilized successfully in the past to examine relationships between writing quality and text features (Crossley, Allen, Snow, & McNamara, 2016; Crossley et al., 2013; McNamara et al., 2019).

2.3. The Current Study

The current study uses computational linguistics and machine-learning models to identify linguistic signatures in writing that predict individual differences in creative and verbal fluency. Specifically, participants took part in a within-subjects writing task that asked them to write out their stream of consciousness and a more structured diary entry. Two traditional cognitive tasks, the AUT and the animal naming task, were used to measure creative and verbal fluency, respectively. We calculated linguistic features along six dimensions of language including descriptive indices, lexical indices related to psycholinguistic word information, syntactic indices related to text complexity, sentiment indices related to valence, and cohesion indices that represented local and global cohesion. We specifically addressed two primary research questions:

1. Overall, how well can all our linguistic features (across all linguistic dimensions) predict individual differences in creative and verbal fluency?
2. How well can linguistic features per dimension (each of six linguistic dimensions) predict individual differences in creative and verbal fluency?

3. Methods

3.1. Participants

This study was approved by a United States Institutional Review Board (IRB) at a large northeastern U.S. institution (IRB #S2021-4). The National Center for Education Statistics lists the undergraduates at this university as primarily full-time students under the age of 24 who identify as white. All participants provided written consent prior to beginning the procedure. Seventy-five undergraduate students were recruited through the university's subject pool for course credit in an introductory psychology course. Five participants were removed for not following the written instructions (N=70). The average age of participants in our sample was 19.59 years old (SD=1.88). Over half of our participants reported being first-year students (54.29%), a little less than a third reported being second-year students (27.14%), with a handful of third-year students (7.14%) and fourth-year students (11.43%). Additionally, over 95% of our sample self-identified as Caucasian/white (95.71%), about 3% identified as Hispanic/Latinx (2.86%), and about 4% identified as Asian/Pacific Islander (4.29%).

3.2. Measures

3.2.1. Creative Fluency

The current study collected data from the AUT to assess creative fluency. This common cognitive task acts as a popular assessment of divergent thinking. Subjects were given one minute to type as many unique or creative alternative uses for a *paperclip* as possible (Harrington, 1975). The number of responses was used to capture creative fluency, an easier dimension to assess creativity (i.e., as opposed to subjective ratings of originality or flexibility that would require more complex assessments such as hand-coding responses or increasing the computational load necessary).

Subjects were instructed to separate each of their responses with a comma to clearly demarcate idea units. For the AUT, responses were qualitatively reviewed to determine appropriateness for inclusion. Responses were excluded if they a) did not list any specific uses (e.g., "metal clip," "found in home office," continuing writing task instead), b) repeated standard, non-creative uses more than once (e.g., "holding paper," "clipping paper"), c) paraphrased a previous response without introducing meaningful variation (e.g., "lockpick," "picking a lock"), or d) included meta-comments (e.g., "idk," "how many creative options are there?”). Responses were scored with relatively liberal inclusion criteria — they did not need to be mutually exclusive (e.g., “pin,” “clothes pin,” “hair pin”) or necessarily practical (e.g., “windchimes,” “dull tattoo needle”). Nevertheless, most retained responses exhibited features typically associated with creativity (e.g., “to test whether a pastry is done baking,” “bra strap connector”).

3.2.2. Verbal Fluency

Verbal fluency was measured using the animal naming task. This task assesses one’s semantic fluency, the ability to verbally produce items within the same semantic category (e.g., types of animals). Instructions asked participants to type as many animals as they could within one minute (Tombaugh et al., 1999).

For the animal naming task, subjects were again asked to separate responses with commas. The first author reviewed the responses to remove any non-animal entries. In cases of uncertainty (such as rare or scientific species names) responses were verified through an online search to confirm their validity as animal names. All levels of animal specificity were accepted as valid (e.g., dog, golden retriever, puppy).

3.2.3. Interrater Reliability

Interrater reliability was evaluated using two-way random-effects models with absolute agreement with the “irr” R package (Gamer et al., 2019). Two raters, independently, scored both fluency measures for each of the 70 participants using only the instructions discussed above. For creative fluency, the intraclass correlation coefficient was ICC(A,1) = 0.989, 95% CI [0.983, 0.993], $F(69, 70) = 188$, $p < 0.001$, indicating high agreement. For verbal fluency, reliability was also high: ICC(A,1) = 0.994, 95% CI [0.991, 0.996], $F(69, 70) = 188$, $p < 0.001$.

3.3. Experimental Procedure

Subjects completed two writing tasks on PsychoPy, an experiment-building service (Peirce et al., 2019) that involved intrapersonal communication: sharing their spontaneous thoughts and completing a diary entry. Instructions for the spontaneous thought stream task were adopted from Sripada and Taxali (2020) and stated: *“For the thought stream task, you will have 10 minutes to generate a stream of your thoughts. This stream of thoughts can encompass any thought that passes into your conscious awareness, including ideas, images, and memories that come to mind. For example, imagine you are on a bus wide awake riding alone late at night. There is little to see and nothing to do. What thoughts would come to mind?”* Similarly, a computerized version of a diary entry task was used for participants to engage in reflective writing (Munezero et al., 2013): *“You will first have 10 minutes to write a diary entry encompassing your current thoughts and ideas. Reflective writing (as in diary writing) encourages you to let go and explore your thoughts and emotions, including your evaluation and perceptions of your thoughts and feelings. For instance, if you were to sit down at night and write in your own diary, what might you say?”*

Participants had ten minutes to respond to each writing task, the order of which was counterbalanced between-subjects. Following the writing tasks, participants completed a series of individual difference measures. Subjects completed the AUT and animal naming task, respectively, before being manually redirected from PsychoPy to Qualtrics. Finally, in Qualtrics, participants completed a series of demographic questions asking subjects about their race, class standing, majors/minors, and age.

3.4. Computational Procedure

The linguistic dimensions of student writing responses were calculated using WAT. Of the thousands of NLP indices automatically produced using this tool, we hand selected five to eight specific indices to represent each major linguistic dimension of interest (i.e., descriptive, cohesion, lexical, lexical diversity, sentiment, syntax). Table 1 shows the descriptions of all linguistic indices (labelled, for reproducibility, with the exact index name used by WAT) that comprise each dimension in the present study. *Descriptive* indices relay information about the length of responses at multiple levels (i.e., word-level, sentence-level). *Cohesion* indices convey information about the amount of overlap within text represented at either the local (i.e., within-sentence) or global (i.e., within-paragraph) level, and can be calculated explicitly (i.e., exact lemma overlap) or semantically (i.e., synonym overlap, word2vec overlap). *Lexical* indices provide insights into the psycholinguistic information about words (e.g., word familiarity, word concreteness). *Lexical diversity* indices are related to how diverse the words used within a text are, consisting mainly of type-token ratio indices (and other slight variations of this formula). *Sentiment* indices identify the valence and semantic properties of words used in the text (e.g., action verb words, familial relationship words). Lastly, *syntactic* indices aid in assessing text complexity, specifically through measures of phrasal complexity, syntactic sophistication, and causal complexity. For each of these dimensions, we chose indices with long-established theoretical roots within each that have been previously linked to other individual differences such as vocabulary knowledge, writing skill, or domain knowledge. Additionally, because the current study is focused on interpersonal writing in general and we don’t expect any major task differences, we aggregated all NLP indices (i.e., across both writing tasks).

Table 1. Linguistic Index Descriptions

Linguistic Dimension	Index	Index Description
Descriptive	Ncontent_words	The number of content words used
	Word_length	The average number of characters per word
	Word_length_variability	The standard deviation of characters per word
	Number_sentences	The number of sentences used
	Average_sentence_length	The average number of words per sentence
	Sentence_length_variability	The standard deviation of words per sentence
Cohesion	Adjacent_overlap_cw_sent_div_seg	The number of content lemma types that occur at least once in the next sentence
	Adjacent_overlap_cw_para_div_seg	The number of content lemma types that occur at least once in the next paragraph
	Syn_overlap_sent_noun	The average sentence to sentence overlap of noun synonyms
	Syn_overlap_sent_verb	The average sentence to sentence overlap of verb synonyms
	Syn_overlap_para_noun	The average paragraph to paragraph overlap of noun synonyms
	Syn_overlap_para_verb	The average paragraph to paragraph overlap of verb synonyms
	Word2vec_1_all_sent	The average word2vec semantic similarity score between all adjacent sentences
	Word2vec_1_all_para	The average word2vec semantic similarity score between all adjacent paragraphs
Lexical	USF_CW	The average contextual distinctiveness score from free association stimuli elicited from content words
	Content_poly	The number of content words in text with a polysemy score (words that have multiple semantic meanings)
	SUBTLEXus_Freq_CW_Log	The average (log transformed) word frequency of content words
	MRC_Familiarity_CW	The mean familiarity score for content words
	MRC_Concreteness_CW	The mean concreteness score for content words
Lexical Diversity	Root_ttr_cw	The number of content word types divided by the square root of the number of content word tokens; Also referred to as "Guiraud's Index"
	Mattr50_cw	The moving average type-token ratio for content words (50-word window)
	Msttr50_cw	The mean segmental type-token ratio for content words (50-word non-overlapping segments)
	Mtld_ma_bi_cw	Version of MTLD (content words only) that takes a moving-average approach to calculating the index. The final value is the mean score when MTLD is run forwards and backwards (but partial factors are not calculated)
	Mtld_ma_wrap_cw	Version of MTLD (content words only) that takes a moving-average approach to calculating the index. The final factor is calculated by wrapping back to the beginning of the text
Sentiment	Vader_negative	Average negative valence of words
	Vader_neutral	Average neutral valence of words
	Vader_positive	Average positive valence of words
	Action_component	Average use of action verbs used from the General Inquirer corpus
	Self_GI_neg_3	Average use of personal pronouns used from the General Inquirer corpus (negations accounted for)
Syntax	Acad_lemma_attested	Percentage of lemmas from the COCA Academic corpus
	Cl_ndeps_std_dev	Clause variety; Standard deviation of the dependents per clause
	Nominal_deps_stdev	Noun phrase variety; Standard deviation of the dependents per nominal
	Cl_av_deps	Clause complexity; Number of dependents per clause
	Av_nominal_deps	Noun phrase complexity; Number of dependents per nominal

We then performed machine-learning models to assess how well these indices combined (i.e., across all linguistic dimensions) could model differences in student fluency scores. Indices were checked for multicollinearity and then mean-centre scaled along with the fluency scores. Multiple machine learning models were trained to predict each fluency score from the combined indices of all linguistic dimensions (all indices listed in Table 1). Eight different algorithms were implemented using the *caret* package in R: linear regression, K-nearest neighbours, linear support vector machine (SVM), polynomial SVM, radial SVM, bidirectional recurrent neural network (BRNN), gradient boosting tree, and Bayesian lasso (Kuhn, 2008). All models were evaluated using 10-fold cross validation and repeated three times until each instance served as the test set. Model success was measured using R^2 and the error rate was assessed using root-mean-square error values (RMSE).

It should be noted that we intentionally implemented multiple machine learning algorithms that varied in complexity and inductive bias — i.e., from linear models to more flexible nonlinear models — to evaluate the robustness and generalizability of our findings. Given the relatively small and complex nature of our dataset, relying on a single modelling approach could have introduced algorithm-specific biases or overfit patterns. By applying the same suite of models across all linguistic feature sets and outcome types (i.e., creative versus verbal fluency), we aimed to assess whether results converged across different modelling paradigms. This triangulation strategy allowed us to examine the stability of predictive patterns and better understand how specific linguistic features contributed to fluency outcomes under differing modelling assumptions. Even when model fits were modest, comparing results across models increased our confidence in the relative informativeness of specific linguistic features and supported more cautious interpretation of model performance.

3.5. Data Availability

The data and code used in this study are publicly available on OSF (https://doi.org/10.21125/edulearn.2024.1725) and did not require or ask about any sensitive information; However, because the writing tasks were unconstrained, many students did disclose personal information. For this reason, we uploaded only the aggregated linguistic data (without including their actual writing samples).

4. Results

4.1. Predicting Individual Differences in Fluencies by Linguistic Feature

On average (and before standardization), students named 7.26 creative uses for a *paperclip* (SD = 3.82) and 17.49 animals (SD = 5.16). Creative and verbal fluency scores were moderately, positively correlated (r=0.35, p<0.001). Additionally, both measures were correlated with several of the linguistic features (listed in Table 2). Most notably, creative fluency was more positively correlated with descriptive and cohesion indices than verbal fluency scores. This finding indicates that while increases in either fluency score are associated with increases in the amount of language produced (descriptive indices), this relationship is strongest for creative fluency, which additionally appears to be linked to having more connected ideas (cohesion indices).

Table 2. Linguistic Index Correlations to Fluency Scores

Linguistic Dimension	Index	Creative Fluency *r*	Verbal Fluency *r*
Descriptive	Ncontent_words	**0.43*****	**0.30*****
	Word_length	–0.13	–0.04
	Word_length_variability	–0.07	0.05
	Number_sentences	**0.21***	**0.25****
	Average_sentence_length	0.01	–0.10
	Sentence_length_variability	**0.26****	0.16
Cohesion	Adjacent_overlap_cw_sent_div_seg	0.05	–0.11
	Adjacent_overlap_cw_para_div_seg	**0.38*****	**0.24****
	Syn_overlap_sent_noun	0.06	–0.09
	Syn_overlap_sent_verb	0.06	–0.14
	Syn_overlap_para_noun	**0.30*****	0.14
	Syn_overlap_para_verb	**0.42*****	**0.19***
	Word2vec_1_all_sent	–0.05	–0.07
	Word2vec_1_all_para	**0.22****	**0.23****
Lexical	USF_CW	–0.05	–0.06
	Content_poly	–0.11	–0.11
	SUBTLEXus_Freq_CW_Log	–0.13	0.05
	MRC_Familiarity_CW	–0.08	–0.05
	MRC_Concreteness_CW	0.04	–0.13
Lexical Diversity	Root_ttr_cw	**0.30*****	**0.17***
	Mattr50_cw	0.08	0.07
	Msttr50_cw	0.02	0.09
	Mtld_ma_bi_cw	0.06	0.16
	Mtld_ma_wrap_cw	0.08	0.02
Sentiment	Vader_negative	0.14	0.09
	Vader_neutral	–0.14	–0.14
	Vader_positive	0.07	0.11
	Action_component	–0.05	–0.15
	Self_GI_neg_3	0.07	0.10
Syntax	Acad_lemma_attested	0.05	–0.07
	Cl_ndeps_std_dev	0.14	0.09
	Nominal_deps_stdev	–0.06	–0.06
	Cl_av_deps	–0.03	0.01
	Av_nominal_deps	0.01	–0.02

Note: $*p < .05$; $**p < .01$; $***p < .001$

Combined model results (see Table 3; best models per DV are bolded) indicate that, overall, linguistic indices can better model differences in student creative fluency (33.8% of variance accounted for) than verbal fluency (23.9% of variance accounted for). Therefore, linguistic indices were able to account for more than a third of the variance in modelling creative fluency (using the linear SVM algorithm) but only about 24% of the variance in verbal fluency (using the BRNN algorithm). These findings indicate that the linguistic properties of one's writing more closely relate to divergent thinking processes that underlie creativity than the convergent thinking processes that underlie verbal fluency. Thus, while creative fluency can be well-captured through NLP indices, the construct of verbal fluency appears to rely more on other cognitive processes (e.g., semantic recall) that depend less on linguistics.

Table 3. Combined Model Results

	Creative Fluency		Verbal Fluency	
Model	RMSE	R^2	RMSE	R^2
Linear Regression	1.221	0.264	1.643	0.161
K-Nearest Neighbours	1.002	0.263	1.045	0.193
Linear SVM	0.970	0.312	1.548	0.160
Polynomial SVM	0.895	0.346	0.979	0.234
Radial SVM	0.949	0.201	0.987	0.166
BRNN	0.916	0.276	1.031	0.239
Gradient Boosting Tree	0.902	0.306	1.174	0.235
Bayesian Lasso	0.932	0.273	0.954	0.200

4.2. Predicting Individual Differences in Creative and Verbal Fluency per Dimension

We next examined how each of our six linguistic dimensions (i.e., descriptive, cohesion, lexical, lexical diversity, sentiment, syntax) could separately predict each of our two fluency scores (i.e., creative fluency, verbal fluency). Results for these models are presented in Tables 4–9.

4.2.1. Descriptive Models

Our descriptive models produced the highest R^2 for both types of fluencies, even higher than that of the combined model (including all linguistic indices across six dimensions). The highest performing algorithm to predict creativity was the Bayesian lasso, while the gradient boosting tree performed best for verbal fluency (see Table 4). Correlations (see Table 2) indicated that the number of content words and number of sentences written were strongly, positively related to both creative and verbal fluencies. However, only with creative fluency did sentence length variability appear to have a significant positive relationship. These results suggest that basic descriptive indices are crucial to modelling fluency scores (both creative and verbal) and reflect a more general ability to just produce written output. The unique relationship between sentence length variability and creative fluency emphasizes the importance of flexibility and complexity in creativity. Thus, creative fluency, specifically, appears to manifest in more diverse and varied language use, distinguishing it from more general verbal fluency abilities.

Table 4. Descriptive Model Results

	Creative Fluency		Verbal Fluency	
Model	RMSE	R^2	RMSE	R^2
Linear Regression	0.870	0.360	0.941	0.275
K-Nearest Neighbours	0.939	0.230	1.002	0.225
Linear SVM	0.861	0.349	0.915	0.283
Polynomial SVM	0.875	0.334	0.931	0.290
Radial SVM	0.917	0.234	0.991	0.221
BRNN	0.879	0.342	0.921	0.277
Gradient Boosting Tree	0.915	0.333	0.939	0.330
Bayesian Lasso	0.942	0.326	0.959	0.296

4.2.2. Cohesion Models

Cohesion models produced results about as successful as the combined model for both types of fluency scores. Models using the Bayesian lasso algorithm performed best in predicting both creative and verbal fluency (see Table 5). Correlations (see Table 2) indicated that many cohesion indices are strongly, positively related to both fluency scores, albeit much more strongly to creativity. For example, the amount of explicit content word overlap, noun/verb synonym overlap, and semantic word2vec overlap among adjacent paragraphs were all much more strongly correlated with creativity than verbal fluency. Interestingly, only more global indices of cohesion (overlap measured across paragraphs) were found to be correlated with fluency scores, as opposed to local indices measured across adjacent sentences in which no significant relationship was found. These results indicate that global cohesion indices play a key role in creative fluency, reflecting the ability to connect and integrate broad ideas. Thus, creative fluency appears to manifest in higher-order organizational processes, such as creating a sense of *cohesiveness* in writing.

Table 5. Cohesion Model Results

	Creative Fluency		Verbal Fluency	
Model	RMSE	R^2	RMSE	R^2
Linear Regression	0.936	0.269	1.081	0.213
K-Nearest Neighbours	0.952	0.228	1.036	0.217
Linear SVM	0.945	0.229	1.057	0.118
Polynomial SVM	0.946	0.274	0.985	0.261
Radial SVM	1.013	0.192	1.030	0.215
BRNN	0.906	0.263	0.943	0.209
Gradient Boosting Tree	1.012	0.198	0.980	0.280
Bayesian Lasso	**0.921**	**0.286**	**0.967**	**0.282**

4.2.3. Lexical Models

For our lexical models, while the best model predicting verbal fluency performed better than our combined model, the best model predicting creativity performed much worse. The best model predicting creative fluency used the linear SVM algorithm and the best predicting verbal fluency used the gradient boosting tree algorithm (see Table 6). However, correlations (see Table 2) indicated no significant relationship between any of the lexical indices for either creative or verbal fluency. These results suggest that both creative and verbal fluencies are minimally related to lexical, word-based properties (e.g., concreteness, frequency). Additionally, the poor performance in predicting creative fluency from lexical indices indicates that creativity relies more on higher-level linguistic dimensions (e.g., cohesion, lexical diversity) than lower-level dimensions such as purely lexical properties.

Table 6. Lexical Model Results

	Creative Fluency		Verbal Fluency	
Model	RMSE	R^2	RMSE	R^2
Linear Regression	1.030	0.159	1.037	0.127
K-Nearest Neighbours	0.996	0.214	1.099	0.161
Linear SVM	1.028	0.147	1.093	0.190
Polynomial SVM	1.018	0.221	1.117	0.140
Radial SVM	1.018	0.170	0.996	0.190
BRNN	1.020	0.175	1.002	0.103
Gradient Boosting Tree	**1.232**	**0.226**	**1.303**	**0.257**
Bayesian Lasso	0.995	0.157	0.967	0.177

4.2.4. Lexical Diversity Models

Like our descriptive models, lexical diversity models performed better than our combined models, although only to a slight degree. The best models predicting both creative fluency and verbal fluency used the polynomial SVM algorithm (see Table 7). Correlations (see Table 2) indicated that only the overall content word type-token ratio index was positively related to our fluency scores. While its correlation with verbal fluency was relatively weak, its correlation with creative fluency was incredibly strong. These findings highlight the significant role that lexical diversity, specifically the type-token ratio, plays in modelling creative fluency. Therefore, using a more diverse range of words appears to be a key feature of creativity. In contrast, the weaker correlation between type-token ratio and verbal fluency suggests that verbal fluency tasks rely more on generating content within a limited lexical scope.

Table 7. Lexical Diversity Models

	Creative Fluency		Verbal Fluency	
Model	RMSE	R^2	RMSE	R^2
Linear Regression	0.889	0.276	0.988	0.149
K-Nearest Neighbours	0.940	0.274	1.014	0.134
Linear SVM	0.884	0.306	0.986	0.264
Polynomial SVM	**0.859**	**0.330**	**0.959**	**0.272**
Radial SVM	0.878	0.311	0.974	0.188
BRNN	0.872	0.300	0.985	0.211
Gradient Boosting Tree	0.913	0.296	1.103	0.198
Bayesian Lasso	0.960	0.292	0.975	0.256

4.2.5. Sentiment Models

Our sentiment models predicting fluency scores performed worse than our combined models. For creative fluency, the model using the radial SVM algorithm performed best. For verbal fluency, the model using the Bayesian lasso algorithm performed best (see Table 8). Correlations (see Table 2) indicated that only the proportion of negatively valenced words was weakly, positively correlated with creative fluency. No sentiment indices were found to be correlated with verbal fluency. Overall, these results suggest that sentiment indices are not strong predictors of either creative or verbal fluency. The poor performance for these sentiment models indicates that fluency scores may be better captured using structural and lexical linguistic dimensions.

Table 8. Sentiment Model Results

	Creative Fluency		Verbal Fluency	
Model	RMSE	R^2	RMSE	R^2
Linear Regression	1.051	0.155	1.012	0.113
K-Nearest Neighbours	1.108	0.296	0.977	0.201
Linear SVM	1.023	0.061	1.008	0.106
Polynomial SVM	1.042	0.201	0.989	0.185
Radial SVM	**1.148**	**0.361**	0.996	0.142
BRNN	1.001	0.148	1.002	0.138
Gradient Boosting Tree	1.394	0.319	1.161	0.184
Bayesian Lasso	0.965	0.180	**0.969**	**0.229**

4.2.6. Syntax Models

Syntax models predicted creative and verbal fluency almost as well as our combined models. The best algorithm to predict creative fluency was K-nearest neighbours, while the best to predict verbal fluency was gradient boosting tree (see Table 9). However, correlations (see Table 2) indicated that no syntactic indices were significantly related to either fluency score. These results indicate that while structural complexity in writing (i.e., syntax) is not linearly correlated with either fluency score, nonlinear algorithms offer increased predictive value. The best models used nonlinear algorithms that captured aspects of fluency scores that linear relationships miss.

Table 9. Syntax Model Results

	Creative Fluency		Verbal Fluency	
Model	RMSE	R^2	RMSE	R^2
Linear Regression	1.030	0.201	1.047	0.174
K-Nearest Neighbours	**0.940**	**0.293**	1.045	0.141
Linear SVM	1.026	0.145	1.083	0.201
Polynomial SVM	0.970	0.204	1.004	0.213
Radial SVM	0.977	0.193	1.104	0.151
BRNN	1.007	0.146	0.993	0.182
Gradient Boosting Tree	1.051	0.243	**1.212**	**0.217**
Bayesian Lasso	0.974	0.133	0.989	0.173

5. Discussion

The current study advances our understanding of how individual differences in cognitive abilities can be computationally analyzed using linguistic features present in students' intrapersonal writing. Various linguistic dimensions were assessed from their writing samples using WAT and the features calculated were used to model differences in creative and verbal fluency. Looking at each linguistic dimension separately, as well as combined, allowed us to investigate the predictive value of each dimension (and the specific indices within them) on the individual differences of interest (i.e., creative and verbal fluency). Broadly, our results suggest that computational linguistics and nonlinear modelling techniques would be a fruitful avenue for learning analytics to implement and predict certain cognitive abilities or individual differences within them, to differing extents.

In our combined models, creative fluency was found to be better predicted from linguistic indices (i.e., accounted for over a third of the model variance) than verbal fluency scores (i.e., which accounted for only about a quarter of model variance). Specifically, key predictors of both creative and verbal fluency included descriptive (e.g., number of content words, sentence length variability) and cohesive indices (e.g., overlap across paragraphs), although these relationships were much stronger with creative, rather than verbal, fluency. Additionally, creative fluency was much more positively correlated with the proportion of unique content words than verbal fluency. This finding highlights the unique role that lexical diversity plays in creativity, with the proportion of unique content words reflecting the writers' varied, or more flexible, language use.

Our findings additionally align with prior studies that emphasize important theoretical differences that underpin divergent and convergent thinking processes. For example, descriptive features of language that tend to be length-based in nature, were very strong predictors of both fluency types, reinforcing their role in the sheer amount of language or ideas produced. Additionally, divergent thinking tasks (i.e., measuring creative fluency) that involve generating novel and unique ideas were best reflected by the degree of linguistic diversity found, such as content word type-token ratios. A final example is that verbal fluency, which aligns more with the idea of convergent thinking (i.e., thinking inside the box), tends to result in weaker relationships to lexical diversity and global cohesion features (Dumas et al., 2021; Gerver et al., 2023). Overall, our results suggest that computational linguistics and nonlinear modelling may be particularly well-suited to measure the multidimensionality of creative expression through writing. Writing can continue to be analyzed as both a creative process and product, representing a window into the complex cognitive processes that occur concurrently (Allen & McNamara, 2015; Crossley, Allen, Snow, & McNamara, 2016).

5.1. Broader Implications

While both creativity and verbal fluency have been tied to writing quality in the prior research, this study provides novel insight into the differing effects of these individual differences in relation to the final written product. First, these results further emphasize the role of multidimensional analyses to uncover complex relationships that may be hidden if expressed in a linear manner only. Thus, the use of NLP and nonlinear machine learning allows for cognitive researchers to better understand how traditional task-based measures of creativity relate to real-world expressions of creativity. Therefore, caution should be exerted when examining linear relationships (i.e., correlations) between creative fluency and other complex processes (e.g., writing quality, reading comprehension).

More broadly, the current study has major implications for the future of learning analytics and educational practices. For example, this work can be utilized to advance stealth assessments of creativity using authentic writing tasks. Techniques such as this could be implemented in automated writing evaluation (AWE) systems to identify individual learners' strengths and weaknesses and provide personalized feedback, explaining the specific ways in which they can improve their writing. This would allow for individual differences to be quietly inferred from the linguistic features used during naturalistic learning activities (i.e., writing) instead of wasting valuable classroom time explicitly assessing every important cognitive skill students posit. Insights may also be able to aid educators and writing researchers in creating ideal constraints that don't just promote writing production, but also creative thinking. Doing so would further provide scalable ways to measure and foster creativity, among other 21st-century skills, within educational contexts.

Overall, our findings have meaningful implications for integrating individualized feedback into AWE systems and intelligent tutoring systems (ITS; Allen et al., 2018; Deane & Zhang, 2020; Shermis et al., 2010). In such systems, student writing could serve as a form of stealth assessment (Öncel et al., 2021; Shute, 2011) to infer individual differences in creative and verbal fluency. For instance, if a student's writing demonstrates limited cohesion or lexical diversity, the system could dynamically recommend targeted activities to support idea generation and language use. A student struggling with cohesion might benefit from tasks that promote idea connectivity, such as the Alternate Uses Task (AUT), whereas a student with low lexical diversity might benefit from semantic retrieval exercises like the animal naming task. Writing cohesion and grammar interventions have long been found to support students by providing personalized feedback. For example, the Writing Pal platform has provided individualized instruction for in-platform games or activities to strengthen the writing skills identified as needing further development (Butterfuss et al., 2022; Crossley et al., 2013 Crossley, Allen, & McNamara, 2016; Roscoe et al., 2018). These kinds of adaptive supports would allow writing platforms to not only assess student writing more formatively but also to guide them toward specific interventions that enhance the cognitive and linguistic skills underlying written fluency.

5.2. Limitations and Future Work

The current study, while offering important insights for educational researchers, is limited in some respects. For example, the participants are undergraduate students specifically seeking out novel educational experiences and may come from higher socioeconomic households. Additionally, we focused primarily on the context of intrapersonal writing for its informal nature, but future work could expand beyond this style of writing and explore other creative writing mediums (e.g., poetry, lyrics, texts). Therefore, in future work researchers should replicate our findings across a more generalized population, using formal writing tasks (e.g., essays), and exploring other forms of creative expression beyond writing.

Additionally, while our linear and nonlinear models identified predictive relationships between linguistic features and fluency scores, we acknowledge that these findings do not imply causality. The observed associations suggest meaningful patterns, but further research is needed to determine whether specific linguistic features actively support or drive fluency. Future experimental work could investigate causal mechanisms, such as whether targeted interventions that promote lexical diversity or cohesion lead to measurable gains in creative or verbal fluency. Additionally, because our study used open-ended, unstructured writing tasks, it remains an open question as to how these linguistic predictors function in more constrained academic writing contexts, such as source-based essays or argumentative writing. Exploring these genre-specific manifestations of creativity would help clarify the generalizability of our findings to real-world educational settings.

Further, our results highlight the potential of computational linguistics in predicting fluency scores. However, future research should combine our techniques with other innovative approaches such as the use of LLMs. Doing so would eliminate the need for subjective human raters in the scoring of creativity and instead utilize pre-trained models, hopefully decreasing scoring biases as well as the time needed to assign scores.

Future research can also incorporate other discourse features into models to improve the accuracy of predicting individual student differences. For example, the rate at which students write/type, eye gaze patterns, and nonverbal gesturing may additionally aid in models predicting creativity. While this study is one of the first to look in-depth at how traditional creativity assessment manifests within real-world forms of creativity, we focus exclusively on creative fluency. However, fluency is only one of four dimensions along which creativity can be assessed (the others being originality, elaboration, and flexibility). Thus, future work should explore how these other dimensions of creative assessment relate to linguistic features.

5.3. Conclusion

In summary, the current study demonstrated how individual differences in cognitive abilities can be successfully analyzed through computational linguistics and nonlinear modelling techniques in the context of educational writing. Findings show how the multidimensional linguistic dimensions expressed through writing have varied predictive strengths that can be leveraged to disentangle effects from individual differences (i.e., creative and verbal fluency). Moreover, this work bridges traditional cognitive measures of creativity and verbal fluency with real-world applications of creativity (i.e., writing). Integrating learning analytics with computational linguistic tools could additionally be used to collect stealth assessments or provide personalized feedback tailored to a student's individual skills and abilities.

Declaration of Conflicting Interest

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

Funding

The authors declared no financial support for the research, authorship, and/or publication of this article.

References

Allen, L. K., Likens, A. D., & McNamara, D. S. (2018). A multi-dimensional analysis of writing flexibility in an automated writing evaluation system. In A. Pardo, K. Bartimote-Aufflick, G. Lynch, S. Buckingham Shum, R. Ferguson, A. Merceron, & X. Ochoa (Eds.), *LAK '18: Proceedings of the 8th international conference on learning analytics & knowledge* (pp. 380–388). ACM Press. https://doi.org/10.1145/3170358.3170404

Allen, L. K., & McNamara, D. S. (2015). You are your words: Modeling students' vocabulary knowledge with natural language processing. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, M. Desmarais (Eds.), *Proceedings of the 8th international conference on educational data mining* (pp. 258–265). International Educational Data Mining Society. https://www.educationaldatamining.org/EDM2015/proceedings/edm2015_proceedings.pdf

Batey, M. (2012). The measurement of creativity: From definitional consensus to the introduction of a new heuristic framework. *Creativity Research Journal, 24*(1), 55–65. https://doi.org/10.1080/10400419.2012.649181

Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., Fink, A., Qiu, J., Kwapil, T. R., Kane, M. J., & Silvia, P. J. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proceedings of the National Academy of Sciences, 115*(5), 1087–1092. https://doi.org/10.1073/pnas.1713532115

Bereiter, C., & Scardamalia, M. (Eds.). (1987). *The psychology of written composition*. Routledge. https://doi.org/10.4324/9780203812310

Bower, J., Acar, S., & Kursuncu, U. (2023). Measuring creativity in academic writing: An analysis of essays in advanced placement language and composition. *Journal of Advanced Academics, 34*(3–4), 183–214. https://doi.org/10.1177/1932202X231211633

Butterfuss, R., Roscoe, R. D., Allen, L. K., McCarthy, K. S., & McNamara, D. S. (2022). Strategy uptake in Writing Pal: Adaptive feedback and instruction. *Journal of Educational Computing Research*, *60*(3), 696–721. https://doi.org/10.1177/07356331211045304

Buczak, P., Huang, H., Forthmann, B., & Doebler, P. (2023). The machines take over: A comparison of various supervised learning approaches for automated scoring of divergent thinking tasks. *The Journal of Creative Behavior*, *57*(1), 17–36. https://doi.org/10.1002/jocb.559

Carey, L. J., & Flower, L. (1989). Foundations for creativity in the writing process: Rhetorical representations of ill-defined options. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 283–303). Springer New York. https://doi.org/10.1007/978-1-4757-5356-1_17
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407
Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125*(2), 159–180. https://doi.org/10.1037/0096-3445.125.2.159
Crossley, S. A., Varner, L. K., Roscoe, R. D., & McNamara, D. S. (2013). Using automated indices of cohesion to evaluate an intelligent tutoring system and an automated writing evaluation system. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial intelligence in education: 16th international conference, AIED 2013, Memphis, TN, USA, July 9–13, 2013, proceedings* (pp. 269–278). Springer. https://doi.org/10.1007/978-3-642-39112-5_28
Crossley, S. A., Allen, L. K., & McNamara, D. S. (2014). A multi-dimensional analysis of essay writing: What linguistic features tell us about situational parameters and the effects of language functions on judgments of quality. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. 197–238). John Benjamins Publishing Company. https://doi.org/10.1075/scl.60.07cro
Crossley, S. A., Allen, L. K., Snow, E. L., & McNamara, D. S. (2016). Incorporating learning characteristics into automatic essay scoring models: What individual differences and linguistic features tell us about writing quality. *Journal of Educational Data Mining, 8*(2), 1–19. https://doi.org/10.5281/ZENODO.3554593
Crossley, S. A., Allen, L. K., & McNamara, D. S. (2016). The writing pal: A writing strategy tutor. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 204–224). Routledge. https://doi.org/10.4324/9781315647500
Davis, G. A. (2004). *Creativity is forever* (5th ed.). Kendall Hunt Publishing.
Deane, P., & Zhang, M. (2020). Automated writing process analysis. In D. Yan, A. A. Rupp, & P. W. Foltz (Eds.), *Handbook of automated scoring: Theory into practice*. Chapman and Hall/CRC. https://doi.org/10.1201/9781351264808
Dumas, D., Organisciak, P., & Doherty, M. (2021). Measuring divergent thinking originality with human raters and text-mining models: A psychometric comparison of methods. *Psychology of Aesthetics, Creativity, and the Arts, 15*(4), 645–663. https://doi.org/10.1037/aca0000319
Dumas, D., & Strickland, A. (2018). From book to bludgeon: A closer look at unsolicited malevolent responses on the alternate uses task. *Creativity Research Journal, 30*(4), 439–450.
Dygert, S. K. C., & Jarosz, A. F. (2020). Individual differences in creative cognition. *Journal of Experimental Psychology: General, 149*(7), 1249–1274. https://doi.org/10.1037/xge0000713
Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity, 5*(3), 178–186. https://doi.org/10.1207/s15327884mca0503_3
Fang, Y., Roscoe, R. D., & McNamara, D. S. (2023). Artificial intelligence-based assessment in education. In B. du Boulay, A. Mitrovic, & K. Yacef (Eds.), *Handbook of artificial intelligence in education* (pp. 485–504). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781800375413/book-part-9781800375413-33.xml
Fredericks, A. D. (2005). *The complete idiot's guide to success as a teacher*. Alpha.
Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *irr: Various coefficients of interrater reliability and agreement* (Version 0.84.1) [Computer software]. https://cran.r-project.org/web/packages/irr/index.html
Gerver, C. R., Griffin, J. W., Dennis, N. A., & Beaty, R. E. (2023). Memory and creativity: A meta-analytic examination of the relationship between memory systems and creative cognition. *Psychonomic Bulletin & Review, 30*(6), 2116–2154. https://doi.org/10.3758/s13423-023-02303-4
Grey, S., & Morris, P. (2024). Capturing the spark: PISA, twenty-first century skills and the reconstruction of creativity. *Globalisation, Societies and Education, 22*(2), 156–171. https://doi.org/10.1080/14767724.2022.2100981
Guilford, J. P. (1950). *Fundamental statistics in psychology and education* (2nd ed.). McGraw-Hill.
Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.
Guilford, J. P., Christensen, P. R., Merrifield, P. R., & Wilson, R. C. (1960). Manual: Alternate uses. Sheridan Supply Co. https://www.mindgarden.com/alternate-uses/78-altu-manual.html#horizontalTab1
Harrington, D. M. (1975). Effects of explicit instructions to "be creative" on the psychological meaning of divergent thinking test scores. *Journal of Personality, 43*(3), 434–454. https://doi.org/10.1111/j.1467-6494.1975.tb00715.x
Hudson, L. (2019). *Frames of mind: Ability, perception and self-perception in the arts and sciences.* Routledge.
Jauk, E., Benedek, M., & Neubauer, A. C. (2014). The road to creative achievement: A latent variable model of ability and personality predictors. *European Journal of Personality, 28*(1), 95–105. https://doi.org/10.1002/per.1941
Kounios, J., Osman, A. M., & Meyer, D. E. (1987). Structure and process in semantic memory: New evidence based on speed–accuracy decomposition. *Journal of Experimental Psychology: General, 116*(1), 3–25. https://doi.org/10.1037/0096-3445.116.1.3

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Marek, R. J., & Ben-Porath, Y. S. (2017). Using the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) in behavioral medicine settings. In M. E. Maruish (Ed.), *Handbook of psychological assessment in primary care settings* (2nd ed., pp. 631–662). Routledge. https://doi.org/10.4324/9781315658407

McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology*, *86*(2), 256–266. https://doi.org/10.1037/0022-0663.86.2.256

McNamara, D. S., Allen, L. K., & Crossley, S. (2019). WAT: Writing assessment tool. In S. Hsiao, J. Cunningham, K. S. McCarthy, G. Lynch, N. Hoover, C. Brooks, R. Ferguson, & U. Hoppe (Eds.), *Companion proceedings of the 9th international conference on learning analytics and knowledge conference (LAK '19)* (p. 256). SoLAR.

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review, 69*(3), 220–232. https://doi.org/10.1037/h0048850

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition, 15*(3), 238–246. https://doi.org/10.3758/BF03197722

Munezero, M., Montero, C. S., Mozgovoy, M., & Sutinen, E. (2013). Exploiting sentiment analysis to track emotions in students' learning diaries. *Proceedings of the 13th Koli Calling International Conference on Computing Education Research*, 145–152. https://doi.org/10.1145/2526968.2526984

Newell, A., & Simon, H. A. (1972). *Human problem solvin*g. Prentice-Hall.

OECD. (2024). *PISA 2022 results: Creative minds, creative schools* (Vol. III). OECD Publishing. https://doi.org/10.1787/765ee8c2-en

Öncel, P., Flynn, L. E., Sonia, A. N., Barker, K. E., Lindsay, G. C., McClure, C. M., McNamara, D. S., & Allen, L. K. (2021). Automatic student writing evaluation: Investigating the impact of individual differences on source-based writing. In M. Scheffel, N. Dowell, S. Joksimović, & G. Siemens (Eds.), *LAK21: 11th international learning analytics and knowledge conference* (pp. 620–625). ACM Press. https://doi.org/10.1145/3448139.3448207

Organisciak, P., Acar, S., Dumas, D., & Berthiaume, K. (2023). Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity*, *49*, Article 101356. https://doi.org/10.1016/j.tsc.2023.101356

Ovington, L. A., Saliba, A. J., Moran, C. C., Goldring, J., & MacDonald, J. B. (2018). Do people really have insights in the shower? The when, where and who of the aha! moment. *The Journal of Creative Behavior*, *52*(1), 21–34. https://doi.org/10.1002/jocb.126

Pei, Z., Zheng, C., Zhang, M., & Liu, F. (2017). Critical thinking and argumentative writing: Inspecting the association among EFL learners in China. *English Language Teaching, 10*(10), 31–42. https://doi.org/10.5539/elt.v10n10p31

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 48–73). Cambridge University Press. https://doi.org/10.1017/CBO9780511763205.005

Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relationships of personality to differential conceptions of creativity: A systematic review. *Psychology of Aesthetics, Creativity, and the Arts, 11*(1), 59–68. https://doi.org/10.1037/aca0000079

Richards, R., Kinney, D. K., Benet, M., & Merzel, A. P. (1988). Assessing everyday creativity: Characteristics of the lifetime creativity scales and validation with three large samples. *Journal of Personality and Social Psychology, 54*(3), 476–485. https://doi.org/10.1037/0022-3514.54.3.476

Rogaten, J., & Moneta, G. B. (2016). Creativity in higher education: The use of creative cognition in studying. In G. B. Moneta & J. Rogaten (Eds.), *Psychology of creativity: Cognitive, emotional, and social processes* (pp. 3–20). Nova Science Publishers.

Roscoe, R. D., Allen, L. K., & McNamara, D. S. (2018). Contrasting writing practice formats in a writing strategy tutoring system. *Journal of Educational Computing Research*, *57*(3), 723–754. https://doi.org/10.1177/0735633118763429

Ruesch, J., & Bateson, G. (2008). *Communication: The social matrix of psychiatry*. Routledge.

Runco, M. A. (1995). Insight for creativity, expression for impact. *Creativity Research Journal, 8*(4), 377–390. https://doi.org/10.1207/s15326934crj0804_4

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Runco, M. A., Millar, G., Acar, S., & Cramond, B. (2010). Torrance tests of creative thinking as predictors of personal and public achievement: A fifty-year follow-up. *Creativity Research Journal, 22*(4), 361–368. https://doi.org/10.1080/10400419.2010.523393

Runco, M. A., Noble, E. P., Reiter-Palmon, R., Acar, S., Ritchie, T., & Yurkovich, J. M. (2011). The genetic basis of creativity and ideational fluency. *Creativity Research Journal, 23*(4), 376–380. https://doi.org/10.1080/10400419.2011.621859

Runco, M. A., & Pritzker, S. R. (Eds.). (2011). *Encyclopedia of creativity* (2nd ed). Academic Press.

Said-Metwaly, S., Van den Noortgate, W., & Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. *Creativity, 4*(2), 238–275. https://doi.org/10.1515/ctra-2017-0013

Sarikaya, I. (2024). Are students with high creativity skills successful writers? A parallel mediation model. *Thinking Skills and Creativity*, Article 101654. https://doi.org/10.1016/j.tsc.2024.101654

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, Article 772. https://doi.org/10.3389/fpsyg.2014.00772

Shermis, M. D., Burstein, J., Higgins, D., & Zechner, K. (2010). Automated essay scoring: Writing assessment and instruction. In P. Peterson, E. Baker, & B. McGraw (Eds.), *International encyclopedia of education* (3rd ed., 20–26). https://doi.org/10.1016/B978-0-08-044894-7.00233-5

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Information Age Publishers.

Simon, H. A. (1985). Some computer models of human learning. In M. Shafto (Ed.), *How we know* (pp. 89-120). Harper & Row.

Smith, R. W., & Kounios, J. (1996). Sudden insight: All-or-none processing revealed by speed–accuracy decomposition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1443–1462. https://doi.org/10.1037/0278-7393.22.6.1443

Sripada, C., & Taxali, A. (2020). Structure in the stream of consciousness: Evidence from a verbalized thought protocol and automated text analytic methods. *Consciousness and Cognition, 85*, 103007. https://doi.org/10.1016/j.concog.2020.103007

Stevenson, C., Smal, I., Baas, M., Dahrendorf, M., Grasman, R., Tanis, C., Scheurs, E., Sleiffer, D., & van der Maas, H. (2020). *Automated AUT scoring using a big data variant of the consensual assessment technique* [Technical report]. Modeling Creativity Project, Universiteit van Amsterdam. https://hdl.handle.net/11245.1/13ad004a-1b61-45a0-8a9a-56d7a165d7ef

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology, 14*(2), 167–177. https://doi.org/10.1016/S0887-6177(97)00095-4

Torrance, E. P. (1972). Predictive validity of the Torrance tests of creative thinking. *The Journal of Creative Behavior, 6*(4), 236–262. https://doi.org/10.1002/j.2162-6057.1972.tb00936.x

Treffinger, D. J. (1986). Research on creativity. *Gifted Child Quarterly, 30*(1), 15–19. https://doi.org/10.1177/001698628603000103

Wati, N. S. (2019). The effect of creativity toward students' achievement in writing ability. *Pedagogy: Journal of English Language Teaching, 6*(2), 141–147. https://doi.org/10.32332/pedagogy.v6i2.1330

Zedelius, C. M., Mills, C., & Schooler, J. W. (2019). Beyond subjective judgments: Predicting evaluations of creative writing from computational linguistic features. *Behavior Research Methods, 51*(2), 879–894. https://doi.org/10.3758/s13428-018-1137-1