Behavior Research Methods (2019) 51:879–894
https://doi.org/10.3758/s13428-018-1137-1

# Beyond subjective judgments: Predicting evaluations of creative writing from computational linguistic features

Claire M. Zedelius[1] · Caitlin Mills[2] · Jonathan W. Schooler[1]

Published online: 27 September 2018


**Abstract**
The question of how to evaluate creativity in the context of creative writing has been a subject of ongoing discussion. A key question is whether something as elusive as creativity can be evaluated in a systematic way that goes beyond subjective judgments. To answer this question, we tested whether human evaluations of the creativity of short stories can be predicted by: (1) established measures of creativity and (2) computerized linguistic analyses of the stories. We conducted two studies, in which college students (with and without interest and experience in creative writing) wrote short stories based on a writing prompt. Independent raters (six in Study 1, five in Study 2) assessed the stories using an evaluation rubric specifically designed to assess aspects of creativity, on which they showed high interrater reliability. We provide evidence of convergent validity, in that the rubric evaluations correlated with established creativity measures, including measures of divergent thinking, associative fluency, and self-reported creative behavior and achievements. Linguistic properties of the short stories were analyzed with two computerized text analysis tools: Coh-Metrix, which analyzes aspects of text cohesion and readability, and Linguistic Inquiry and Word Count, which identifies meaningful psychological categories of the text content. Linguistic features predicted the human ratings of creativity to a significant degree. These results provide novel evidence that creative writing can be evaluated reliably and in a systematic way that captures objective features of the text. The results further establish our evaluation rubric as a useful tool to assess creative writing.

**Keywords** Creativity · Creative writing · Language · Coh-Metrix · Linguistic inquiry and word count (LIWC)

Creative writing is a form of expression that affords great freedom. Writers are constrained only by the limits of their imagination (and sometimes a deadline). Writing is also very personal. Good writers are thought to develop an authentic "voice" or writing style that makes their work unique and recognizable among others. Confirming this common perception, research has shown that students in a writing class were able to identify both famous authors and even their peers from samples of their writing (Gabora, O'Connor, & Ranjan, 2012). This uniqueness of expression is an aspect of creative writing that is highly valued, but it poses a challenge when it comes to evaluating the creativity of a piece of writing. If every piece of writing is unique, can we nonetheless compare different pieces and evaluate them according to some objective standard, or is creativity purely subjective? We addressed this question across two studies by testing whether the objective characteristics of writers' language usage (assessed via computerized linguistic analyses) predicted how creative human judges would find their writing (assessed via rubric scoring). We also examined to what extent human evaluations of creative writing correlated with established individual difference measures of creativity, including idea generation, associative thinking, and real-life creative or artistic behaviors and achievements. The studies established the evaluation rubric as a useful and reliable approach for assessing creativity in the domain of creative writing.

✉ Claire M. Zedelius
claire.zedelius@gmail.com

[1] University of California, Santa Barbara, CA, USA

[2] University of British Columbia, Vancouver, British Columbia, Canada

## Evaluating creativity: Is it all subjective?

Creativity has been defined in various ways (see, e.g., Amabile, 1983b; Boden, 1994; Guilford, 1967). A key element of many definitions is that creative products are novel, surprising, or original (e.g., Barron, 1955; Boden, 1994;

Rhodes, 1961). In addition, definitions also commonly stress usefulness or appropriateness as important aspects of creativity (see Kampylis & Valtanen, 2010; Mumford, 2003; Runco, 1988; Sternberg & Lubart, 1999). The question of how to evaluate creativity in the context of creative writing has been subject to ongoing discussion (e.g., Blomer, 2011; Kantor, 1975; Kaufman, Plucker, & Baer, 2008; May, 2007; Newman, 2007). Carl Rogers (1954), speaking more generally about all creative acts (including but not limited to creative writing), went so far as to say that we often cannot even recognize a creative product as such, because "the very essence of the creative is its novelty, and hence we have no standard by which to judge it" (p. 252). Thus, according to Rogers, attempts to systematically compare and evaluate a piece of writing for their creativity might potentially miss the truly creative aspects of the work.

Among those who think that we can evaluate the creativity of a piece of writing, existing approaches can be arranged on a continuum ranging from a focus on subjective judgments to a greater focus on objective evaluation criteria. On one end of the continuum are scholars who maintain that creative writing can only be evaluated on the basis of subjective liking or gut feelings (Kantor, 1975; Newman, 2007)—a valid position, though one that poses obvious problems for the scientific study of creativity. A more empirically oriented perspective has been proposed by Amabile (1983a, 1996), who developed the "consensual assessment technique" (CAT). This technique assigns the evaluation process to expert judges. In the domain of creative writing, this includes experienced writers, editors, and teachers. The experts are asked to make holistic evaluations of creative works according to their own idiosyncratic standards. The judges work independently and without direction. Although this technique inherently relies on subjective evaluations, the idea behind the approach is that experts base their subjective evaluations on a deeper knowledge and understanding of the field. Studies using this technique both in the domain of creative writing and for evaluating other artwork have shown that expert juries typically show impressively high agreement between raters (Kaufman & Baer, 2012; Kaufman, Baer, Cole, & Sexton, 2008; Kaufman, Gentile, & Baer, 2005; Swim, Borgida, Maruyama, & Myers, 1989). These findings have helped establish the "gold standard" status of the CAT for evaluating creativity (Carson, 2006; Kaufman, Baer, & Cole, 2009). But because the CAT relies on subjective evaluations based on idiosyncratic standards—standards that typically are not explicitly articulated—its effectiveness can ultimately only be judged by the level of agreement between raters.

A different approach to evaluating creativity consists of using a rubric. Rubrics typically consist of ordinal or interval scales (for instance, from poor to excellent) to systematically score various aspects of creative works based on explicitly defined criteria. This approach is frequently used to evaluate the quality of creative writing, particularly in educational settings (see Andrade, 2005; Arter & McTighe, 2000; Mozaffari, 2013; Reddy & Andrade, 2010; Rodriguez, 2008). Many of the rubrics developed for these purposes include aspects of writing not necessarily related to creativity, such as spelling, grammar, punctuation, syntax, and textual organization. Few rubrics, however, have been developed specifically to assess novelty, originality, or other creative aspects of writing (see Mozaffari, 2013; Young, 2009).

Evaluation rubrics have the practical advantage of not requiring expert judges. Students can potentially even learn to use a rubric to evaluate their own writing. Another advantage of using a rubric over the CAT or similar measures is that the criteria used to determine what is considered creative are transparent and consistent, allowing students to have some insight into how their writing will be judged and how they can improve it. The rubric criteria as an assessment of creativity remain subjective. That is, judges may disagree about whether a rubric measures what *they* personally consider creative. However, a key assumption of the rubric approach is that judgments about the degree to which a particular piece of writing fulfills these criteria are based on more objective characteristics of the text. Of course, the effectiveness of this approach depends on the degree to which this assumption is true. Unfortunately, this issue has received little attention by researchers. As was observed by Rezaei and Lovorn (2010),

> the reliability and validity of holistic writing assessment strategies have been studied for years; however, analytical rubric-based assessment has not been adequately and experimentally studied. Although analytic rubrics have emerged as one of the most popular assessment tools in progressive educational programs, there is an unfortunate dearth of information in the literature quantifying the actual effectiveness of the rubric as an assessment tool. (p. 21)

The present research addressed this issue. Our goal was to test whether human evaluations of short stories made with a creativity evaluation rubric (1) overlap with a variety of established creativity metrics and (2) can be predicted from easily quantifiable computerized metrics of the writers' language usage, and thus, from characteristics inherent in the texts. If successful, this would speak to the usefulness of using evaluation rubrics—and our rubric in particular—to measure creativity in the context of creative writing. It could also contribute to the development of future tools that might be able to evaluate creativity on the basis of automated text analyses. We conducted two studies using two participant samples: (1) psychology undergraduates, who were not recruited on the basis of any particular interest or training in creative writing, and (2) a broader sample of students (including English and other majors) who had an interest in creative writing, and some of

whom reported to be aspiring writers. We asked the students to write a short story based on a writing prompt, and to complete a number of established creativity measures that assess individual differences in divergent thinking, associative fluency, and self-reported creative and artistic behavior and achievements. We then had independent raters evaluate the creativity of the short stories using a novel rubric that we adapted from an existing rubric designed to assess different aspects of creativity (Mozaffari, 2013). In our evaluations of the short stories, we defined creativity more narrowly than the terms *novelty*, *originality*, and *usefulness*. As we will discuss below, the rubric included originality as one component of creativity, but it also assessed other criteria specific to creative writing. In addition to assessing interrater reliability, we validated the rubric by examining correlations between the story evaluations and the other creativity measures. This served to provide evidence that the rubric indeed, as intended, captures aspects of writing that are related to what is commonly considered creativity. We then conducted computerized analyses of the linguistic characteristics of the texts.

## Automated text analyses to evaluate writing quality

Although limited work has been done to study creative writing using automated, computationally derived linguistic tools, the previous research has highlighted the potential predictive power of such tools (McNamara, Crossley, Roscoe, Allen, & Dai, 2015; Newman, Pennebaker, Berry, & Richards, 2003; Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014; Varner, Roscoe, & McNamara, 2013). Two of the most commonly applied text analysis tools are Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004) and Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2001, 2007; Tausczik & Pennebaker, 2010).

Coh-Metrix produces automated measures of cohesion and readability by analyzing multiple hierarchical levels of a text. These go beyond surface level features of the text to tap causal, temporal, and intention cohesion. Cohesion reflects the characteristics of the text that contribute to the construction of a situation model (i.e., overall mental representation). These characteristics range from words to phrases to sentences that help build a coherent text representation. The metrics produced by Coh-Metrix are derived through a combination of latent semantic analysis (LSA), syntactic parsers, and traditional difficulty indices like word and sentence length. Coh-Metrix was attractive for the present research because it has been validated across several studies (see McNamara & Graesser, 2012; McNamara, Louwerse, McCarthy, & Graesser, 2010) and has been shown to successfully predict human ratings of essay quality (Crossley, Roscoe, & McNamara, 2011; Graesser et al., 2004; McNamara, Crossley, & McCarthy, 2010)—although, to date, it has not been used to predict human ratings of creativity.

LIWC is another popular text analysis tool that uniquely assesses psychologically meaningful categories, such as emotional content or social relationships (for a review, see Tausczik & Pennebaker, 2010). LIWC analyzes each word in a given text by comparing it to a set of 80 predefined dictionaries (e.g., the word "happy" as an example from the positive emotion dictionary), then computes the percentage of words in the text that fall in each category. Importantly, LIWC has been found to successfully predict psychologically relevant outcomes, including deception (Newman et al., 2003), students' grades (Robinson, Navea, & Ickes, 2013), and college academic success (Pennebaker et al., 2014). For example, how many first-person pronouns students used in a self-introduction essay for an undergraduate course predicted the students' final grade at the end of the course (Robinson et al., 2013).

Previous studies have shown that Coh-Metrix and LIWC can be used to predict human ratings of student writing. These studies found that human ratings of "good" writing can be predicted by the sophistication of the language, as judged by linguistic indices such as lexical diversity, word frequency, and syntactic complexity (Crossley & McNamara, 2011a, 2011b, 2012; Crossley et al., 2011; McNamara, Crossley et al., 2010). However, these studies focused exclusively on predicting noncreative aspects of writing quality. The qualities that are appreciated in good *creative* writing seem more elusive and are not necessarily expressed through sophisticated language. Hemingway, for instance, is famous for his plain and simple writing style (e.g., Charyulu, 2016). Thus, it was an open question whether Coh-Metrix and LIWC would be useful in the context of creative storytelling.

Only one previous study (Kim, Lee, & Lee, 2012) has explored the possible relationship between students' performance on a creativity task (divergent thinking) and their language usage, using the Korean version of LIWC (K-LIWC). However, it is important to note that the researchers did not try to predict the creativity of *the writing itself* from the writers' language usage. Instead, they were interested in the relationship between students' language usage in a writing task and their creativity in an unrelated task measuring divergent thinking (the alternate-uses task, which is frequently used to measure creative ability; see Guilford, 1967; Torrance, 2008). Moreover, the students were asked to engage in "stream-of-consciousness writing"—that is, to write down the spontaneous thoughts that went through their head, with no explicit goal related to creativity. The authors found that 40 of the 84 K-LIWC variables correlated with students' scores on the divergent-thinking task. These results were purely exploratory and have not been replicated, so we do not put much weight on any one particular correlation, especially since there is an elevated risk of Type I errors with this high number of tests (C.

H. Lee, Kim, Lim, & Lee, 2015). However, the findings are some of the only available evidence suggesting more globally that individual differences in performance on creativity tasks may be reflected in a person's language usage. It remains unclear, though, whether the correlations are driven by aspects of the writing that make the writing creative, or by aspects that reflect other personal characteristics.

## Creative writing rubric and key features of interest

We used Coh-Metrix and LIWC indices (see the Method section for indices description) to predict three evaluative aspects of students' creative stories that have been proposed by Mozaffari (2013) to constitute the qualities of creative language: Image, Voice, and Originality. *Image* assesses the degree to which the writing evokes vivid mental images in the reader. This includes visual images, but also sounds, smells, physical sensations, and emotions. Writing high in Image can transport the reader into the story and/or the mind or the body of a character. It is achieved by avoiding generalizations, abstractions, and other elements of flat writing and by providing rich and concrete descriptions of events (e.g., "I tilted forward, and there I was, horizontal and flying, the cool air against my face. I flapped my arms a little, like a baby bird, and I ascended. Soon I was about 50 feet off the ground, in full flight. I put my arms out in front of me and gently circled the dry, yellow cornfield," rather than "I stretched my arms out and began to fly.").

*Voice* assesses the degree to which the author has succeeded in creating his or her own unique, recognizable voice, or writing style. This can be achieved through a variety of stylistic tools, including a unique and particular choice of words (e.g., rare, old-fashioned, or slang words, words that are untypical in the particular context of the story, or self-created words or expressions; e.g., "the buttermilky fog," "her don't-give-a-damn clothes"), use of interesting or sophisticated sentence structures and punctuation, use of metaphors and other tropes, or through a unique perspective or attitude of the narrator (e.g., sarcasm, wittiness, darkness).

*Originality* assesses to what degree the story idea or plotline is original, that is, unlike other stories. The comparison here is both with stories from other study participants and with other familiar stories (e.g., from popular books, plays, comic books, movies, TV shows, musicals, tales, and urban legends). Thus, this criterion can be said to capture the novelty aspect common in many definitions of creativity (Barron, 1955; Boden, 1994; Rhodes, 1961), whereas the other criteria capture other aspects more specific to the domain of writing.

These three aspects were measured with a rubric that we adapted from an existing rubric developed by Mozaffari (2013).[1] Mozaffari's rubric was designed with the specific goal of capturing aspects of writing that make a story creative (rather than merely "good writing"). The rubric addressed the shortcomings of previous rubrics, which contained criteria that were either too general or vague (e.g., story is original), or irrelevant to creativity (e.g., correct grammar and spelling). However, Mozaffari's rubric has not been used or further refined by independent researchers, raising questions about its reliability and ease of scoring, and its applicability to different texts. We made small modifications to the rubric that served to make it more easily applicable. Specifically, we edited the descriptions of the evaluation criteria to make them less ambiguous (our raters initially had problems understanding the scoring criteria), and we slightly modified the rating scale.

Mozaffari's (2013) rubric also had not been validated with other creativity measures. Thus, the degree to which the rubric succeeded in capturing aspects of creativity is unknown. Our research addressed this evidence gap by testing to what extent evaluations of participants' short stories with our modified rubric correlate with established measures of creativity that could reasonably be expected to correlate with creative writing skills.

Before writing the short stories, participants completed the following creativity measures: (1) the alternate-uses task, a frequently used measure of a divergent thinking (Guilford, 1967; Torrance, 2008). In this task, participants are asked to come up with as many unusual and creative uses as possible for common objects. Because the task requires original thinking, it is conceivable that greater performance on the task would be associated with creative writing skills, and with developing a unique voice. (2) Compound remote associates problems, which are verbal problems designed to measure associative fluency (Gabora, 2010; Mednick, 1962). Because the problems converge on a single solution, they are also often regarded as a measure of convergent creativity (Bowers, Regehr, Balthazard, & Parker, 1990; Zedelius & Schooler, 2015). Performance on these problems has also been found to correlate with intelligence and, in particular, verbal ability (e.g., C. S. Lee, Huggins, & Therriault, 2014). For these reasons, they can be expected to correlate with creative writing ability. (3) Finally, participants answered a questionnaire assessing their real-life creative behaviors and achievements (Dollinger, 2003). Such behaviors and achievements may reflect creative ability and/or an interest in the arts, and may thus be related to creative writing. In sum, we predicted that participants' scores on these creativity measures would correlate moderately with their short story evaluations from the rubric scoring.

[1] Mozaffari's (2013) rubric included one additional dimension of creative writing: Characterization. We found that the short format of the stories used in this study did not allow for deep enough character development to reliably evaluate this aspect. This dimension should be more applicable to longer texts, however.

## Method

### Participants and sample size determination

We expected that the correlations between the rubric scores and the other creativity measures and linguistic features would range from small to medium effect sizes, on the basis of previous studies predicting rubric-based assessments from computational linguistic features (Crossley & McNamara, 2011a). Based on Harris's (1985) formula for determining minimal sample size (described in Van Voorhis & Morgan, 2007), regression analyses with up to seven predictors would require an absolute minimum of 70 participants. To be able detect correlations with small effect size, we aimed for a higher number, and terminated data collection at the end of the academic quarter.

The participants in Study 1 were 133 undergraduate students (88 female, 44 male, and one who did not indicate gender; mean age = 19.3 years, *SD* = 1.4). All of the participants majored in psychology and participated in the study in exchange for course credit.

The participants in Study 2 were 128 undergraduate students (83 female, 44 male, and one who did not indicate gender; mean age = 20.1 years, *SD* = 3.8). The students participated in exchange for money. The major difference in Study 2 was that we made a specific effort to recruit a more heterogeneous sample of participants, including students with an interest in creative writing. We recruited the participants through e-mails and flyers targeting students in the university's English department and the College of Creative Studies, who are interested in creative writing. Th eparticipants' majors were in English, communication, comparative literature, psychology, film and media, "undecided," or other fields (including fields not directly related to creativity or writing). Of the participants in Study 2, when asked whether they would describe themselves as an aspiring writer, 30.5% responded "yes," 27.3% responded "no," and 42.2% responded "not sure." Moreover, when asked how many hours they spent on writing or writing-related activities (e.g., editing or research for a writing project) in an average week, 18.8% responded "none–less than an hour," 28.9% "1–2 hours," 35.9% "3–5 hours," 9.4% "6–8 hours," 6.3% "9–11 hours," and 0.8% "12 hours or more." There was no overlap between participants across the two studies.

### Materials and procedure

Participants performed a battery of creativity measures, in counterbalanced order, followed by a creative writing assignment.

**Alternate uses task (AUT)** Participants were asked to generate unusual or creative uses for two objects; a tin can and a cardboard box (order counterbalanced). They were given 90 s for each object. Responses were scored by independent raters for fluency (number of valid uses) and originality (a more subjective assessment of how rare or creative each use was); uses that were unusual but not specific to the given object (e.g., "throw it into the ocean") were not considered original.

**Compound remote associates (CRA) problems** CRA problems are verbal problems often used as a measure of associative fluency (Bowers et al., 1990; Mednick, 1962). On each trial, participants are shown three words (e.g., "board, magic, death"), and are asked to find a fourth word that can be combined with each word in the set to form a compound word or phrase (the solution in this example is "black," as in "black board," "black magic," "black death"). Because the solution word is a *remote* (i.e., weak) associate of each of the target words, solving the problems requires accessing a wide network of distantly associated words in memory (e.g., Benedek & Neubauer, 2013; Garbora, 2010) while inhibiting strongly associated words that aren't correct solutions (see, e.g., Harkins, 2006).

Participants received 32 CRA problems. For each problem, they were given 30 s to indicate via button press that they had found the solution. After pressing the button, they had 10 s to type in their answer.

**Creative Behavior Inventory (CBI) short form** The CBI short form (Dollinger, 2003) is a 28-item self-report measure of creative and artistic behaviors and achievements on a 4-point scale ranging from *never did this* to *did this more than five times*. The items describe activities and accomplishments in the visual, literary, and performing arts and crafts. Cronbach's alpha in the present studies was .90 in both Studies 1 and 2.

**Creative writing assignment** The creative writing assignment was given at the end of the study session. Participants were given 20 min to write a short fictional story on the computer (using the text editing program Microsoft Word). They were instructed to use the following writing prompt as the basis for their story:

> Create a character who has suddenly and unexpectedly attained some sort of power. In the wider perception of the world the level of authority may be small or great, but for this person, the change is dramatic. Write about the moment in which your character truly understands the full extent of his or her newfound power for the first time.

## Coh-Metrix indices

We used the five commonly applied principal components from Coh-Metrix to predict short story evaluations based on the rubric. These components have been found to explain over 50% of the variability in texts across grade levels and text categories (including narrative and informational texts; Graesser, McNamara, & Kulikowich, 2011): (1) *Narrativity* reflects how well the text aligns with the narrative genre by conveying a story, procedure, or sequence of actions and events. Texts higher in narrativity are more conversational and more "story-like". (2) *Deep cohesion* is computed on the basis of the degree to which different ideas in the text are cohesively tied together with connective words signifying causality or intentionality. (3) *Referential cohesion* reflects how well content words and ideas are connected to each other across the entire text. Texts higher in referential cohesion have greater overlap between sentences throughout the story with regard to explicit content words and noun phrases. (4) *Syntactic simplicity* is computed on the basis of the simplicity of the syntactic structures in the text. Simple syntax uses fewer words and simpler sentence structures with fewer embedded sentences that require the reader to hold information in working memory. (5) *Word concreteness* is computed on the basis of the degree to which words in the text evoke concrete mental images that are easier to process than abstract words.

## LIWC dimensions

We used LIWC version 2015 to analyze participants' writing. LIWC contains more than 80 different measures of text content and style, which are all highly specific (e.g., swearwords; categories such as, "friend," "home," etc.). LIWC is particularly attractive because some of its produced indices seem theoretically irrelevant to creativity (at least as measured by the rubric). However, using over 80 indices in our analyses was an unnecessary approach, since many indices were theoretically irrelevant to assessing creativity and, in addition, a number of the indices are highly intercorrelated. Therefore, instead of using all 80 + indices as potential predictors of creativity, we opted for a hypothesis-driven approach and selected a small number of LIWC indices that were theoretically relevant to the criteria according to which the rubric's measures were evaluated. We opted for this approach in order to reduce the risk of model overfitting, to avoid Type I errors, and to yield results that could be more readily interpreted. Although the rubric contains some aspects that likely cannot easily be related to specific LIWC indices (e.g., an "unreliable narrator" as a stylistic tool), other aspects (e.g., concrete descriptions of visual, auditory, and sensory details or the use of rare words) may be captured by particular LIWC indices, making the hypothesis-driven approach feasible.

Because writing rich in Image was characterized in the rubric as vividly bringing to live images, sounds, smells, tastes, thoughts, feelings, and bodily sensations, we selected the following five LIWC indices as candidates to capture these specific aspects: (1) *Insight*: This category contains words such as "think," "know," "consider"—words that can be used to describe thoughts, feelings, and internal images ("I imagined opening my arms and leaping off the balcony"). (2) *See*: This contains words such as "view" or "saw," which can describe visual images. (3) *Hear*: This contains words such as "listen" and "hearing," which are relevant to describe sound experiences. (4) *Feel*: This contains words, such as "feels" or "touch," that can describe feelings and bodily sensations (e.g., "she feels a strange tingling sensation"). (5) *Body*: This contains words, such as "cheek" or "hands," that are useful to describe feelings and bodily sensations (e.g., "My mouth was dry, and I felt my knees buckle.").

Writing that scores high in Voice has been described as using stylistic tools such as choice of particular or unusual words (e.g., rare or old-fashioned words or informal words), noteworthy sentence structures (e.g., very short or very long sentences), punctuation (e.g., many hyphens, colons, etc.), and a particular emotional tone or narrative attitude. We predicted that these aspects might be captured by the following six LIWC variables: (1) *Informal language*: This category comprises informal language such as swear words, netspeak ("lol"), and nonfluencies like "er," "umm," relevant to the scoring of Voice. (2) *Dictionary words*: This is the number of words from the LIWC dictionary a text contains. We selected this variable because broad and varied word usage may be useful for creating a unique voice. (3) *Number of words per sentence*, (4) *Punctuation*, and (5) *Use of commas specifically*: These three categories are simple counters of words per sentence, punctuation marks, and commas. They were selected because they appear relevant to creating noteworthy sentence structures or using punctuation for stylistic purposes. (6) *Authenticity*: This variable measures how personal and honest a person's language sounds to listeners. Language scoring high in authenticity contains many first-person words and present-tense verbs, among other features. This variable was selected because authentic language may be indicative of a particular narrative attitude that helps create a recognizable voice.

Originality scores reflected whether a story idea or plot is original, as compared to others. Thus, in contrast to Image and Voice, Originality is defined largely contextually (i.e., original relative to other stories) and is therefore more difficult to relate to specific linguistic features inherent in the text. We therefore chose not to select specific theoretically driven LIWC indicators to predict Originality. We nonetheless conducted exploratory regression analyses with all LIWC indicators included as predictors in order to test whether Originality can be systematically predicted by the LIWC categories. We report those results after the hypothesis-specific results for Image and Voice.

## Other measures

Other questionnaire measures unrelated to creativity were administered as part of another study, and those are not reported here.

## Rubric scoring

The stories were rated by independent raters (six [two men, four women] in Study 1, five [two men, three women], in Study 2), which included the first author and undergraduate students from the Psychology and English departments. None of the raters had been participants in either study. The raters were blind to the identity of the writers (including any demographic or creativity measures) and rated the stories in a pseudorandomized order. Raters were trained in using the adapted rubric to evaluate sample stories for Image, Voice, and Originality using a 3-point scoring system (*poor*, *fair*, *very good*). A 4-point scoring system (as proposed by Mozaffari, 2013) was initially considered, but it became obvious during the training process that the highest scores (3 and 4) could not be reliably distinguished. (The adapted rubric and scoring key can be found in the Appendix.) All raters rated all stories. After having rated 20% of the stories, any noteworthy discrepancies in their ratings were discussed before continuing to rate the remaining stories. The raters' average scores were used as the final measures for analysis purposes. Finally, the raters also made subjective "gut feeling" evaluations, on a scale from 1 (*Do not like at all*) to 10 (*Like very much*). However, in retrospect, because these ratings were made immediately following the rubric evaluations, we do not think they can be interpreted as an independent holistic measure of creativity. We therefore do not report results associated with these ratings.

# Results

## Reliability and validity of the rubric evaluations

First we tested the reliability and convergent validity of the rubric scores. As Tables 1 and 2 show, the three measures (Voice, Image, and Originality) of the rubric are all moderately correlated with each other ($r$ = .30 to .64 in Study 1, $r$ = .41 to .69 in Study 2), indicating that they measure related but distinct aspects of creative writing. We then tested whether the three measures can be assessed with high interrater reliability, a question that was not directly tested in Mozaffari's (2013) original study. We used a fully crossed design in which all raters rated all stories. Hence, in line with Hallgren's (2012) recommendation, we used the average-measures intraclass correlation (ICC) to assess interrater reliability. We found it psychologically meaningful and justifiable to treat the rating scale as an ordinal scale, so that disagreements between raters of greater magnitude (ratings of "poor" vs. "very good") would be penalized more than disagreements of smaller magnitude (e.g., "poor" vs. "fair"). Raters showed excellent reliability. In Study 1, the ICCs were .93 for Image, .92 for Voice, and .90 for Originality. In Study 2, the ICCs were .90 for Image, .91 for Voice, and .86 for Originality.

To assess the convergent validity of the adapted rubric, we tested whether the rubric's measures of creative writing correlated with the other creativity measures. The correlations for both studies are shown in Tables 1 and 2. As can be seen, the correlations were moderate and significant, but for a few exceptions. For the psychology undergraduate sample in Study 1, Image and Voice were correlated with nearly all other measures of creativity (the only exception being the nonsignificant correlation between Image and fluency on the AUT), whereas Originality failed to correlate consistently with the other creativity measures. Among the broader student sample (Study 2), Image, Voice, and Originality showed consistent correlations with performance on the creativity tasks. This is first

**Table 1** Correlations between creative writing rubric measures and other creativity measures in Study 1

| | Story Voice | Story Originality | CBI | AUT Fluency | AUT Originality | RAT Accuracy |
|---|---|---|---|---|---|---|
| Story Image | $r$ = .638<br>$p$ < .001 | $r$ = .303<br>$p$ < .001 | $r$ = .222<br>$p$ = .010 | $r$ = .135<br>$p$ = .120 | $r$ = .245<br>$p$ = .004 | $r$ = .179$p$ = .039 |
| Story Voice | | $r$ = .413<br>$p$ < .001 | $r$ = .226<br>$p$ = .009 | $r$ = .171<br>$p$ = .049 | $r$ = .226<br>$p$ = .009 | $r$ = .223<br>$p$ = .010 |
| Story Originality | | | $r$ = .006<br>$p$ = .948 | $r$ = .110<br>$p$ = .208 | $r$ = .179<br>$p$ = .039 | $r$ = .103<br>$p$ = .248 |
| CBI | | | | $r$ = .141<br>$p$ = .105 | $r$ = – .006<br>$p$ = .945 | $r$ = .197<br>$p$ = .023 |
| AUT Fluency | | | | | $r$ = .627<br>$p$ < .001 | $r$ = .206<br>$p$ = .017 |
| AUT Originality | | | | | | $r$ = .105<br>$p$ = .229 |

**Table 2** Correlations between creative writing rubric measures and other creativity measures in Study 2

| | Story Voice | Story Originality | CBI | AUT fluency | AUT originality | RAT accuracy |
|---|---|---|---|---|---|---|
| Story Image | $r = .692$<br>$p < .001$ | $r = .413$<br>$p < .001$ | $r = .204$<br>$p = .021$ | $r = .230$<br>$p = .009$ | $r = .242$<br>$p = .006$ | $r = .223$<br>$p = .011$ |
| Story Voice | | $r = .437$<br>$p < .001$ | $r = .098$<br>$p = .732$ | $r = .227$<br>$p = .010$ | $r = .181$<br>$p = .042$ | $r = .317$<br>$p < .001$ |
| Story Originality | | | $r = .031$<br>$p = .731$ | $r = .267$<br>$p = .002$ | $r = .199$<br>$p = .025$ | $r = .262$<br>$p = .003$ |
| CBI | | | | $r = .173$<br>$p = .049$ | $r = .055$<br>$p = .538$ | $r = .176$<br>$p = .044$ |
| AUT Fluency | | | | | $r = .228$<br>$p = .009$ | $r = .236$<br>$p = .007$ |
| AUT Originality | | | | | | $r = .105$<br>$p = .233$ |

evidence that the rubric measures are related to other aspects of a person's creativity, such as their ability at creative idea generation and verbal problem solving and their engagement in the arts and crafts.

## Computerized text analyses

**Coh-Metrix** For analyses using Coh-Metrix, we used the five principal component measures from Coh-Metrix—narrativity, syntaxtic simplicity, word concreteness, referential cohesion, and deep cohesion—as predictors in three separate regression models predicting the three rubric measures—Image, Voice, and Originality. For Image, the total variance explained by the model was 25.1% for Study 1 and 36.5% for Study 2. For Voice, the total variance explained by the model was 40.8% for Study 1 and 24.3% for Study 2. For Originality, the total variance explained by the model was lower: 11.8% for Study 1 and 17.5% for Study 2. Table 3 lists the coefficients for both studies. Referential cohesion emerged as the most consistent predictor across both studies, significantly predicting both Image and Voice. The negative direction of the relationships indicates that less cohesion was predictive of higher creativity scores. Image was also predicted by deep cohesion and word concreteness consistently over both studies. Originality was less consistently predicted by the Coh-Metrix variables: Only syntactic simplicity was significantly related to Originality, and this realtionship was limited to Study 2.

**LIWC** Given the high number of indices produced by LIWC (over 80), a set of hypothesis-driven LIWC measures were used to predict Image (*insight*, *see*, *hear*, *feel*, *body*) and Voice (*informal language*, *dictionary words*, *number of words per sentence*, *punctuation*, *use of commas*, *authenticity*). We conducted separate regression analyses to predict Image and Voice from the specific LIWC variables in each study (four analyses total). The results are shown in Tables 4 and 5. We also conducted two exploratory regression analyses with Originality as the dependent variable (separately for each study), the results of which are reported below.

For Image, the total variance explained was 37.6% for Study 1 and 39.4% for Study 2. Image was consistently and negatively predicted by the category *insight*. It was positively predicted by words from the category *feel*, and it also showed a significant association with *body*-related words in Study 1 and a statistically trending relationship in Study 2. Contrary to our prediction, words related to seeing and hearing did not predict Image scores.

For Voice, the total variance explained by the LIWC categories was 45.4% for Study 1 and 31.5% for Study 2. Voice was consistently positively predicted by language categorized as *authentic*, and negatively predicted by number of words in the LIWC *dictionary* (i.e., fewer dictionary words corresponded to higher Voice scores). We also found an association between punctuation and Voice scores that was significant in Study 1 and trending in Study 2. Informal words predicted Voice scores only in Study 1, and commas only in Study 2. Words per sentence did not predict Voice scores.

For Originality, which was defined largely contextually, we did not select specific LIWC indicators but conducted exploratory regression analyses in which all nonredundant LIWC indicators were included. The overall regression model was not significant for either Study 1 [$F(90, 132) = 1.281$, $p = .188$] or Study 2 [$F(90, 127) = 1.390$, $p = .131$].[2]

## Discussion

Creative writing is a highly valued form of artistic expression. It is also highly valuable for creativity research. As an ecologically valid measure of real-world creativity, it can be a useful

[2] The same exploratory analyses performed with Image and Voice as dependent variables did yield significant overall models [*Image:* Study 1, $F(90, 132) = 3.489$, $p < .001$; Study 2, $F(90, 127) = 2.969$, $p < .001$; *Voice:* Study 1, $F(90, 132) = 3.130$, $p < .001$; Study 2, $F(90, 127) = 2.185$, $p < .005$]..

**Table 3** Coefficients of Coh-Metrix predictors of rubric evaluations in Studies 1 and 2

| | Study 1: Image | Study 2: Image | Study 1: Voice | Study 2: Voice | Study 1: Originality | Study 2: Originality |
|---|---|---|---|---|---|---|
| Narrativity | $\beta = .135$<br>$p = .181$ | $\beta = .153$<br>$p = .139$ | $\beta = .12$<br>$p = .178$ | $\beta = .096$<br>$p = .394$ | $\beta = .033$<br>$p = .759$ | $\beta = -.306$<br>$p = .011$ |
| Syntactic simplicity | $\beta = -.05$<br>$p = .569$ | $\beta = .104$<br>$p = .254$ | $\beta = .053$<br>$p = .498$ | $\beta = .098$<br>$p = .326$ | $\beta = .136$<br>$p = .154$ | $\beta = .226$<br>$p = .033$ |
| Word concreteness | $\beta = .286$<br>$p = .001$ | $\beta = .276$<br>$p = .001$ | $\beta = .014$<br>$p = .857$ | $\beta = -.01$<br>$p = .911$ | $\beta = .148$<br>$p = .115$ | $\beta = -.042p$<br>$= .650$ |
| Referential cohesion | $\beta = -.427$<br>$p < .001$ | $\beta = -.418$<br>$p < .001$ | $\beta = -.577$<br>$p < .001$ | $\beta = -.443$<br>$p = .001$ | $\beta = -.241$<br>$p = .033$ | $\beta = -.01$<br>$p = .938$ |
| Deep cohesion | $\beta = -.234$<br>$p = .003$ | $\beta = -.306$<br>$p < .001$ | $\beta = -.285$<br>$p < .001$ | $\beta = -.101$<br>$p = .222$ | $\beta = -.111$<br>$p = .193$ | $\beta = -.039$<br>$p = .655$ |

alternative or addition to the more artificial tasks and problems typically used to measure aspects of creativity in the lab. Most research that has assessed creative writing has employed the CAT to evaluate creativity. Although this approach has been shown to yield highly valid and reliable evaluations (see Kaufman & Baer, 2012), it comes with some practical concerns, since it relies on willing experts or quasi-experts to participate as judges, sometimes under budget and time constraints. Moreover, although useful as an evaluative tool, it provides little insight into what makes one piece of writing more creative than another. An alternative approach, often used in educational settings and educationally relevant research, is the use of systematic evaluation rubrics. Our goal was to explore whether human judgments of creativity made with an evaluation rubric can be predicted by computerized linguistic analyses of the same stories. This would provide evidence that the rubric indeed captures aspects of creativity inherent in the text itself.

Raters evaluated short stories from student using an evaluation rubric—our own slightly modified version of an existing rubric (Mozaffari, 2013)—specifically designed to assess three aspects of creativity; the extent to which the writing speaks to the imagination (Image), the extent to which the writing has a distinctive voice or style (Voice), and the extent to which a story is original. Our findings showed that the raters attained excellent interrater reliability. The reliability was higher than in Mozaffari's original study, suggesting that our modifications to the original rubric achieved their goal of making the rubric easier to apply. We further extended Mozaffari's work by showing that creative writing scores made with the rubric correlated with individual differences measures of creativity not directly related to the writing task. Specifically, we found that, in both studies, students' scores for Image and Voice correlated with their self-reported creative and artistic behaviors and achievements, and with their performance on the divergent thinking task (particularly originality scores) and associative fluency. Originality scores from the rubric showed somewhat inconsistent relationships with the other creativity measures. Among the broader student sample with an interest in creative writing (Study 2), originality was correlated with performance on the divergent thinking and associative fluency tasks (though not with self-reported creative behavior), but among the group of psychology undergraduates (Study 1), originality scores for the stories were correlated only with originality scores on the divergent thinking task. All in all, these correlations provide first evidence that the rubric measures are related to other aspects of a person's creativity and artistic expression, thus validating the rubric as a tool for assessing creativity. This is important, since the rubric was specifically designed with the

**Table 4** Coefficients of LIWC predictors of Image in Studies 1 and 2

| | Study 1 | Study 2 |
|---|---|---|
| Insight | $\beta = -.224$<br>$p = .003$ | $\beta = -.246$<br>$p = .001$ |
| See | $\beta = -.014$<br>$p = .858$ | $\beta = .097$<br>$p = .225$ |
| Hear | $\beta = .046$<br>$p = .532$ | $\beta = .080$<br>$p = .274$ |
| Feel | $\beta = .176$<br>$p = .032$ | $\beta = .374$<br>$p < .001$ |
| Body | $\beta = .433$<br>$p < .001$ | $\beta = .179$<br>$p = .068$ |

**Table 5** Coefficients of LIWC predictors of Voice in Studies 1 and 2

| | Study 1 | Study 2 |
|---|---|---|
| Authenticity | $\beta = .239$<br>$p = .001$ | $\beta = .250$<br>$p = .001$ |
| Words per sentence | $\beta = -.017$<br>$p = .864$ | $\beta = -.165$<br>$p = .052$ |
| Dictionary | $\beta = -.266$<br>$p < .001$ | $\beta = -.232$<br>$p = .004$ |
| All punctuation | $\beta = .287$<br>$p = .031$ | $\beta = .147$<br>$p = .094$ |
| Comma | $\beta = .090$<br>$p = .324$ | $\beta = .278$<br>$p = .001$ |
| Informal language | $\beta = .234$<br>$p = .004$ | $\beta = .019$<br>$p = .819$ |

goal of identifying writing that is not merely "good," but creative. This is also what distinguishes the rubric from other existing rubrics, which tend to place a greater emphasis on correct spelling and grammar or sophisticated writing (see Andrade, 2005; Young, 2009).

The correlations are also interesting in their own right, in so far as they speak to the domain generality or specificity of creativity and the relationship between creative writing and other creative skills or abilities (Baer, 1991; Baer & Kaufman, 2005; Silvia, Kaufman, & Pretz, 2009). Previous studies have reported mixed evidence for a relationship between students' creative writing ability and performance on other types of creative tasks (Baer, 1991; Kaufman, Evans, & Baer, 2010). The correlations we observed between creative writing evaluations and our other creativity measures are all positive, suggesting some overlap among students' skills at writing in a way that is image-rich, original, and has a unique voice, their ability to generate original ideas and find connections between remotely associated words or concepts, and their engagement in creative or artistic behaviors. The correlations are of small to moderate in size, however, and it is unclear to what extent they may be attributed to shared creative processes underlying these different tasks or other factors, such as motivation, intelligence, personality (e.g., openness), or interest in artistic pursuits (see Wolfradt & Pretz, 2001). Further research will be needed to explore this question further.

After having established reliability and validity of our rubric evaluations of creativity, our main question of interest was whether the rubric scores correlated with linguistic features of the texts, analyzed with the text analysis tools Coh-Metrix and LIWC. This question was of major interest to us because the rubric can only be a valuable alternative to subjective evaluation methods if it is indeed based on objective characteristics of the text. Of the five principal component measures derived from Coh-Metrix—narrativity, syntaxtic simplicity, word concreteness, referential cohesion, and deep cohesion—referential cohesion emerged as a consistent predictor of Image and Voice (but not Originality). Image was also predicted by deep cohesion and word concreteness, again consistently over both studies. Originality, however, was more difficult to predict from linguistic features. Syntactic simplicity predicted Originality in Study 2, but not in Study 1. Given the nonreplication, this finding should be interpreted cautiously and reexamined in future research.

What is perhaps most interesting about these findings is that greater levels of deep cohesion and referential cohesion were predictive of *lower* scores or Image and Voice. Cohesion (achieved through cohesive devices such as reference, conjunction, identifiers of causal relationships or intentionality, and repetition of important content words) is important for building coherent textual representations in the mind of readers, and has been linked to text quality. However, it has previously been found that the relationship between cohesion and text quality is not necessarily straightforward (Crossley & McNamara, 2011). Texts low in cohesion force the reader to make inferences to bridge conceptual gaps and make connections in text that are not explicitly stated (McNamara, 2001; McNamara, Crossley et al., 2010; O'Reilly & McNamara, 2007). This makes low cohesive texts more difficult to comprehend, but only for readers with little background knowledge. It is possible that stories with low cohesion, because they require more work on the part of the reader and stimulate inferences (Graesser et al., 2011), may also be perceived as more creative, since they leave more to the reader's imagination.

In trying to predict creativity scores from the 80 + measures assessed by LIWC, we focused on linguistic features and psychological constructs that we expected to be most relevant to the rubric measures. Because Originality was defined largely contextually (i.e., how unique a plot or story idea was in comparison to the other stories), we did not try to predict Originality from specific hypothesis-driven LIWC categories. We did, however, conduct exploratory analyses with all LIWC indices as predictors, which yielded nonsignificant results for the overall models in both studies. The LIWC measures are meant to capture the percentage of words that fall in a predefine category, which is likely not able to capture a story's originality. Future work trying to predict originality from computational linguistic tools may attempt to compare stories using topic models (Blei, 2012), which could assess how different topics are across essays, for instance.

On the basis of the rubric's scoring criteria for Image (writing high in Image should vividly bring to live images, sounds, smells, tastes, thoughts, feelings, and bodily sensations), we thought that words related to cognitive and sensory processes and bodily experiences would be the most likely predictors. According to our expectations, Image was consistently (over both studies) predicted by words related to bodily sensations and characteristics of the body (the categories *feel* and *body*). Interestingly, words related to sensory processes, specifically seeing and hearing (*see*, *hear*) did not predict Image, suggesting that descriptions of visual and auditory experiences are not as evocative of mental images as descriptions focusing on emotional and bodily sensations. Words describing thought processes (the *insight* category) significantly predicted Image, but negatively so. A possible explanation could be that words describing thought processes (e.g., a character reflecting on what they "think," "know," or "consider" to be true, rather than a character simply experiencing the world) may be rather abstract. The Image criterion is characterized more by concrete descriptions of events (consistent with its positive relationship with word concreteness found in the Coh-Metrix analyses).

For the rubric category voice—for which evaluation criteria highlighted the use of stylistic tools, punctuation, and narrative attitude—we expected that linguistic features (e.g., punctuation, sentence length, used of informal words, and language authenticity) more than psychological content categories would be likely predictors. Authentic language (this is characterized by frequent use of the first person

perspective and present tense and by avoiding relatives or comparison words) indeed positively predicted Voice scores. So did punctuation, although commas specifically and sentence length (words per sentence) each predicted Voice scores only in one of the two studies. Informal language likewise was related to Voice only in the first of the two studies. Thus, these findings need to be interpreted with caution, and should be replicated. Interestingly, the more words from the dictionary a writer used, the lower their score on Voice. This negative relationship could mean that using more unique or very rare words, rather than stereotypical dictionary words, contributes to the perception that the writer created their own voice.

Although our predictions for the LIWC analyses were based on somewhat tentative hypotheses, and not all our findings are replicated in both studies, most findings show a consistent pattern over both studies, despite the different participants samples. Thus, although these findings certainly warrant further replication, they are encouraging in that they suggest that human ratings of creativity can at least to some extent be predicted from objectively measured linguistic features of the texts.

## Limitations

The present research has noteworthy limitations—some related to the broader question of how to measure and quantify creativity, whereas others relate to the specific tools and analyses used here to assess features of creative language.

Many definitions of creativity stress novelty and originality as key aspects of creativity. These aspects are also contained in the way we operationalized creativity here in the context of writing, in so far as novelty is important for creating a unique voice and novelty and originality are integral to coming up with an original idea for a story. Novelty and originality are notoriously difficult to quantify because they are highly context-dependent. Whether a piece of work is deemed new or original depends in part on who judges the piece and how much knowledge they have of the domain in which creativity is evaluated (Kaufman & Baer, 2012). It also depends on the reference group within which a creative product has emerged and is to be evaluated (Silvia et al., 2008). In our research we explicitly defined the Originality criterion in a context-dependent manner. We think that this was the reason Originality was difficult to predict from features inherent in the text. The originality of participants' stories correlated with their performance on independent creativity measures (more so in Study 2), but unlike with Image and Voice, we failed to predict it consistently from the Coh-Metrix components and LIWC indicators. This demonstrates the clear limits of computerized and exclusively text-based evaluation methods of creativity. However, computerized text-based evaluations can still be highly valuable as part of a repertoire of creativity measures, including expert evaluations. In future research, it would therefore be interesting to include the CAT as an alternative evaluation approach and to test the extent to which computerized text analyses overlap with it and differ from it.

Other limitations of our studies concern the specific text analysis methods we used. We selected two frequently used yet methodologically very different text analysis tools: Coh-Metrix for its ability to analyze coherence and cohesion at various hierarchical levels of a text, and LIWC for its focus on psychologically meaningful categories of text content. Coh-Metrix, with its five established principal components, was well suited for conducting data-driven analyses to predict creativity. LIWC, with its 80 + measures of punctuation and content words presented more of a challenge for conducting data-driven, exploratory analyses. Predicting an outcome from that many variables poses an elevated risk of Type I errors (and statistically correcting for multiple comparisons may in turn make it hard to detect small effects). To address this concern, we decided to take a hypothesis-driven approach, selecting specific LIWC predictors that we thought were theoretically relevant to the criteria according to which the rubric's measures were evaluated. This approach still presented us with palpable limits, because a writer has an infinite number of stylistic and other tools at their disposal to bring a scene or a character to live or to create a unique, recognizable voice. Some of the examples mentioned in the rubric could reasonably be expected to relate to particular word usage or punctuation, but others perhaps can't be captured at all at the level of individual word categories. Moreover, the hypotheses we formulated could easily be replaced by various other theoretically plausible hypotheses. Thus, LIWC is perhaps not the ideal tool for assessing creative language. However, given the success of LIWC for predicting other psychologically relevant outcomes (Newman et al., 2003; Pennebaker et al., 2014; Robinson et al., 2013), this should be examined further in future research.

Although we point out that there are clear limitations to evaluating creative writing using automatized linguistic analysis, we want to stress that our goal was not to develop an automatized scoring system that could predict creativity with great certainty or replace human evaluations. At this point, computerized metrics cannot fully comprehend and appreciate a writer's ability to come up with highly novel or beautiful metaphors or a highly captivating story. Moreover, they cannot easily put a text into a broader context, be it a cultural or historical context or a comparison with other stories. The limited results for Originality scores illustrate that point. However, our results are the first one to suggest that something as seemingly subjective and ephemeral as whether a story speaks to the imagination or is written in a strong voice can have some basis in objective features of the text.

## Conclusion

The present results provide novel evidence that creative writing can be evaluated reliably and in a systematic way with an

evaluation rubric that captures aspects of creative language. Over two studies, our results showed that human ratings of creativity made with the help of a rubric can to some extent be predicted from objectively measured linguistic features of the texts. These results establish the evaluation rubric as a useful tool to assess creative writing and demonstrate "proof-of-concept" evidence that at least some aspects of creative writing can be captured by computerized measures—evidence that warrants more attention in follow-up studies. We hope that future research will follow up on these findings. Future work should continue to explore creativity through systematic assessments of creative writing, using a combination of subjective holistic evaluations and reliable and validated evaluation rubrics and linguistic analyses.

**Author note** This research was supported by Grant RFP-15-09 from the Imagination Institute (www.imagination-institute.org), funded by the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Imagination Institute or the John Templeton Foundation.

## Appendix: Evaluation rubric

### Image

- The writing speaks to the imagination. It evokes vivid mental images (including visual images, but also sounds or smells).
- The writing contains rich, **concrete details**.

Example 1: "*I tilted forwards, and there I was, horizontal and flying, the cool air against my face. I flapped my arms a little, like a baby bird, and I ascended. Soon I was about fifty feet off the ground, in full flight. I put my arms out in front of me and gently circled the dry, yellow cornfield.*" Rather than: "*I stretched my arms out and began to fly.*"

Based on our rating experience: Having about 1–3 rich details (depends how rich) tends to make a story a 2, having more than 3 will likely make it a 3 ➔ see Table 6.

- The text **lacks** these elements of flat writing
- **Abstractions:** referring directly to concepts which cannot be experienced through the senses (e.g., "*love,*" "*hopelessness*")
- **Generalizations:** using words that are too vague to be visualized (e.g., "*everything,*" "*her life*") without further elaboration
- The writing uses **literary tropes** such as metaphors, simile, or personification. *E.g.:* "*Those Saturdays sat on my head like a coal scuttle, and if mama was fussing, as she was now, it was like somebody throwing stones at it.*" The sentence uses personification (sitting is a human characteristic which is used for Saturday) and simile (when mother was fussing it was like throwing stone at the head).
- The writing uses thinking or speaking **voices** *if this is applicable to the story*

*e.g.,* "*Following her mother with her eyes, she started to talk: 'May . . . May I . . . May I stay at home tonight? I have to study . . .'*" rather than "*She asked her mother if she could stay home because she had to study.*"

### Examples from the study

- Chloe went to bed that night with a terrible migraine. It was as if an unyielding metal ring was being cranked tighter and tighter around her skull until her brain screamed with pain that went up, down, sideways but never out. [. . .] Walking through the halls of the dorm she feels a strange tingling sensation in her brain, as if a swarm of subtle whispers are bombarding her thoughts. [. . .] Chloe runs past her like a hurricane and into the lecture hall that is already brimming with students. (3)
- I could smell the cigarettes on his breath, which choked the little life out of my pitiful existence. Time had seemed to stop for me then, and I could feel the pain of the many times before this, where the smell of cigarettes invaded my senses, and I choked as a knife was held against my skin. [. . .] his words were seething, and the knife sliced the skin of my neck just in the slightest, sending pain trickling through my body. Blood dripped down my neck, staining the shirt my uncle gave me [. . .] (25)
- I caught a glimpse of a bird. Something about the way it could fly freely, without any responsibility [. . .] I imagined opening my arms and leaping off the balcony, only to dive down and dip back up into flight. [. . .] I lost myself in the cool air that brushed my hair behind my head and tickled my skin. (55)
- There was no other way to describe the pain coursing through my veins, the pounding in my head. My mouth was dry, and I felt my knees buckle into the snow before I could stop them. I placed my palms against the sides of my head and tried to breathe. [. . .] A swirl of orange danced across my palms, lighting my across my fingers. They flickered with the wind but refused to disappear. [. . .] the orange swirls continued to merrily dance. [. . .] A harsh gust of wind blew through the trees, and I realized night was falling. (15)

## Voice

The author has created his/her own unique, recognizable voice. A voice is unique and recognizable if you, when reading another piece by the same author, would recognize that it is written by the same author.

This can be achieved through purposefully (i.e., in a way that makes sense, not randomly) using **stylistic tools (/physical/visible style)** such as:

- particular choice of words, e.g., rare words, words not typically used in that context, old-fashioned words, slang words, interjections (Oh! Wow!); comic-book words ("swoooosh"); self-created words or expressions ("the buttermilky fog," "her don't-give-a-damn clothes"); jargon
- sentence structures, e.g., mixing very short with very long sentences; long, unbroken stream-of-consciousness type sentences; sequential build-up of action ("everything around him stopped: the people froze, his hotdog which flew out of his hand froze, the cars froze, and everything just froze (except Sam).")
- frequent use of metaphors or comparisons ("His alarm screamed at the top of its lungs")
- italic or bold print, all-caps, particular punctuation, paragraph spacing

It can also be achieved by through a **unique perspective or attitude** of the narrator, for instance:

- Humor, sarcasm, wittiness
- Darkness (e.g., written like a horror story); lightheartedness (e.g., like a lovely, innocent, peaceful story for kids)
- Deliberate simplicity or complexity (e.g., "He was a baker. But that day, being a baker was less fun than usual. Baking the bread did not give him pleasure. Kneading the dough made his hands hurt."; "This was exactly the arrant pedantry up with which I would not put!")

oHow do you know if a style is *deliberately* simple and not just poorly written? See if it "works." Is it *actually* easy to comprehend, or are the sentences all confusing and ambiguous (e.g., What does "it" refer to???)? Is the author constantly switching between past and present tense and uses phrases that make no sense? Does the simplicity make sense given the story? *E.g.*, it makes sense if it's sort of a kids story, or told from the perspective of a child or naïve character.

- Mixing what happens in the story with a self-aware reflection on it (e.g., "it's finally happened. [. . .] After all these years [. . .] I finally have a superpower. I mean, it's superhuman . . . that's for sure. An incredibly underwhelming superpower, I have to admit. [. . .] It's weird, I know. But it wasn't always like this. One day I was in the bathroom, doing my business, and I heard someone in the stall next to me [. . .]")
- Unreliable narrator (e.g., Remember the underwater story. It was written rather realistically, not like a fantasy story, and the narrator clearly was convinced of her power which, to an outside observer, seems more like a dream or wishful/desperate fantasy: "She could breathe underwater. She could speak to the dolphins, stingrays, and clown fish. [. . .] until the sunset broke, when she was forced to swim back to shore and return to her bedroom back on earth [. . .] crawl back under the covers and wait for her mother to come get her.")
- Special narrator (e.g., perspective of a very young child or an alien who thinks and speaks in a way that's very different from our way of thinking and speaking)
- Interesting choice of who narrates (e.g., story in form of an obituary, newspaper report, chatroom conversation, text-message conversation, interview, etc.)

Based on our rating experience: Using ***some*** *stylistic tools* **OR** having a unique perspective or attitude of the narrator tends to make a story a 2. Having **both** will likely make it a 3. But having voth is not required for a 3. Using ***many*** *stylistic tools* (without the perspective/attitude) **OR** having a ***very*** unique perspective or attitude of the narrator (without stylistic tools) can each also make it a 3.

### Examples from the study

- "First there was an alien invasion, then there was that weird announcement, then there were the floating rocks in the sky, and finally there was this. [. . .] Out of Reina's limp form a glowing flower was blooming, facing towards Cordelia the way other flowers face the sun. It beckoned her. In some way that Cordelia couldn't understand, she felt compelled towards it, to touch it and see what would happen." (76)
- "It was a pitiful existence, and it led to this day, where blood was meant to leak upon the pavement, and my life was to end. [. . .] I reached out to him, and reaped his soul straight from his body, tearing through his chest with my grim hands, all in a dark mist, which resonated from the ground. The next words came through, from a memory long ago: 'I am death, the destroyer of worlds,' and with that a scythe materialized in a purple mist, and my eyes grew dark. I realized what I had been given, a power to correct the evils of man, to fight like my uncle did long ago. In my death, the reaper was born." (25)
- "Diana had first met Isabel when she moved from Boston to this small town of Fresno and she sat next to her in English. Both immediately hit it off, expressing their weird love for Shakespeare that no one else could ever understand. And from that moment, they became best

friends. To Diana, Shakespeare gave her an escape from her normal boring life and inspired her to write her own short stories, mainly about the dreams that she would have throughout the day. Everything from her prince on a white horse to a ghost chasing her down the hallway, she wrote it out into her daily diary which now was a collection of all of her short stories (or dreams)." (62)

## Originality

The story plot or basic idea is original, that is, it is very much unlike the other stories in the pile, and unlike other familiar stories (e.g., stories you know from books, movies, TV shows, tales and urban myths, etc.).

(A story can be original if it resembles *some* obscure story you've read or seen before but aren't even sure other people have heard of. It's not original if it resembles every other well-known comic book or movie)

Originality can be achieved in different ways, such as:

- The power is an unusual power or a not-so-unusual power gained or executed in an unusual or unexpected way (e.g., unusual context or boundary conditions)
- The author made an unusual or unexpected choice about where the plot would lead (e.g., the story starts out like a kitschy teenage romance but ends in an unexpectedly dark place; the story starts very realistically and ends up being unexpectedly mystical; the story starts out dark but becomes unexpectedly funny)
- The author made an unusual or unexpected choice *in the face of this particular writing prompt*, such as: taking the concept of power literally (having power in the sense of being able to rule over or boss around other people) or interpreting it in some other way that's very different from the typical interpretation
- The characters or setting are very unusual (e.g., animals or aliens as main characters; story set on a different planet)

### Examples from the study

- "What if one day you woke up and discovered that you could snap your fingers and the meal you were thinking of would appear just like that! [. . .] While Amanda could not turn invisible or fly, like she always thought powers were, she was very satisfied and proud of her power because besides getting what she wanted, she was able to help a greater society." (12)
- "No matter how hot he would raise the water to him, that it did not feel hot. [. . .] The once always boiling hot oatmeal that he was used to, now was too cold as well. [. . .] His warmed toned skin by the end of the day now seemed a faint blue. All these changes frightened him. After all he was a man that liked for everything to be normal, expected." (16)
- "I finally have a superpower. I mean, it's superhuman . . . that's for sure. An incredibly underwhelming superpower, I have to admit. It is the ability to know who hasn't washed their hands after using the bathroom." (31)

Each story is rated along all three criteria, using a 1–3 rating scale. Table 6 explains what ratings of 3, 2, and 1 mean.

**Table 6** Summarized scoring criteria

| Criterion | Rating 3 – Very good | Rating 2 – Fair | Rating 1 – Poor |
|---|---|---|---|
| Image | Several rich, concrete details<br>AND/OR<br>Several literary tropes<br>Direct thought/speech (*if applicable*)<br>No or very few abstractions or generalizations | Some moderately rich details<br>AND/OR<br>One literary trope<br>Direct thought/speech (*if applicable*)<br>Some abstractions, generalizations, and judgments | No or very few concrete details<br>No literary tropes<br>No direct thought/speech even though applicable<br>Lots of abstractions, generalizations, and judgments |
| Voice/Style | Either: *[Some* stylistic tools *combined with* a *somewhat* unique perspective or attitude]<br>OR<br>*[Many* stylistic tools<br>OR<br>A *very* unique perspective or attitude] | One or two stylistic tools<br>OR<br>A somewhat unique perspective or attitude | No stylistic tools<br>No unique perspective or attitude |
| Originality | Highly original plot/idea<br>Unlike most other stories in the pile<br>Doesn't seem like a story I've read or seen many times before<br>Very original power or ordinary power in an unusual context | Somewhat original plot/idea<br>On par with many other stories, better than some<br>Seems somewhat familiar; like other stories I may have read or seen<br>Power not necessarily terribly original but not entirely cliché | Unoriginal plot/idea<br>Very much like a lot of other stories in the pile<br>Very familiar, definitely similar to stories I've read or seen<br>Power is cliché with zero original twist |

## References


Amabile, T. M. (1983a). The social psychology of creativity. New York, NY: Springer.

Amabile, T. M. (1983b). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, *45*, 357–376. https://doi.org/10.1037/0022-3514.45.2.357

Amabile, T. M. (1996). Creativity in context. Boulder, CO: Westview Press.

Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, *53*, 27–31.

Arter, J., & McTighe, J. (2000). Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance. Thousand Oaks, CA: Corwin Press.

Baer, J. (1991). Generality of creativity across performance domains. *Creativity Research Journal*, *4*, 23–39.

Baer, J., & Kaufman, J. C. (2005). Bridging generality and specificity: The amusement park theoretical (APT) model of creativity. *Roeper Review*, *27*, 158–163.

Barron, F. (1955). The disposition toward originality. *Journal of Abnormal and Social Psychology*, *51*, 478–485.

Benedek, M., & Neubauer, A. C. (2013). Revisiting Mednick's model on creativity-related differences in associative hierarchies. Evidence for a common path to uncommon thought. *Journal of Creative Behavior*, *47*, 273–289.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*, 77–84.

Blomer, Y. (2011). Assessment in creative writing. *Wascana Review*, *43*, 61–73.

Boden, M. A. (1994). What is creativity? In *Dimensions of creativity* (pp. 75–118). Cambridge, MA: MIT Press.

Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology*, *22*, 72–110.

Carson, J. C. (2006). *Creativity and mental illness.* Invitational panel discussion hosted by Yale's Mind Matters Consortium, New Haven, CT.

Charyulu, G. M. (2016). Stylistic devices in Hemingway's novels: A study on *The Old Man and the Sea*. *Journal of the English Literator Society*, *1*, 306–310.

Crossley, S. A., & McNamara, D. S. (2011a). Text coherence and judgments of essay quality: Models of quality and coherence. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), Expanding the space of cognitive science: Proceedings of the 33rd Annual Meeting of the Cognitive Science Society (pp. 1236–1241). Austin, TX: Cognitive Science Society.

Crossley, S. A., & McNamara, D. S. (2011b). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life Long Learning*, *21*, 170–191.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*, 115–135.

Crossley, S. A., Roscoe, R. D., & McNamara, D. S. (2011). Predicting human scores of essay quality using computational indices of linguistic and textual features. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), Proceedings of the 15th International Conference on Artificial Intelligence in Education (pp. 438–440). Berlin, Germany: Springer.

Dollinger, S. J. (2003). Need for uniqueness, need for cognition, and creativity. *Journal of Creative Behavior*, *37*, 99–116.

Gabora, L. (2010). Revenge of the "neurds": Characterizing creative thought in terms of the structure and dynamics of memory. *Creativity Research Journal*, *22*, 1–13.

Gabora, L., O'Connor, B., & Ranjan, A. (2012). The recognizability of individual creative styles within and across domains. *Psychology of Aesthetics, Creativity, and the Arts*, *6*, 351–360.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*, 223–234.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*, 193–202. https://doi.org/10.3758/BF03195564

Guilford, J. P. (1967). The nature of human intelligence. New York, NY: McGraw-Hill.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*, 23–34.

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation–performance relationship. *Journal of Personality and Social Psychology*, *91*, 436–455. https://doi.org/10.1037/0022-3514.91.3.436

Harris, R. J. (1985). A primer of multivariate statistics (2nd ed.). New York, NY: Academic Press.

Kampylis, P. G., & Valtanen, J. (2010). Redefining creativity—Analyzing definitions, collocations, and consequences. *Journal of Creative Behavior*, *44*, 191–214.

Kantor, K. (1975). Evaluating creative writing: A different ball game. *The English Journal*, *64*, 72–74.

Kaufman, J. C., & Baer, J. (2012). Beyond new and appropriate: Who decides what is creative? *Creativity Research Journal*, *24*, 83–91.

Kaufman, J. C., Baer, J., & Cole, J. C. (2009). Expertise, domains, and the consensual assessment technique. *Journal of Creative Behavior*, *43*, 223–233.

Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the Consensual Assessment Technique. *Creativity Research Journal*, *20*, 171–178.

Kaufman, J. C., Evans, M. L., & Baer, J. (2010). The American Idol effect: Are students good judges of their creativity across domains? *Empirical Studies of the Arts*, *28*, 3–17. https://doi.org/10.2190/EM.28.1.b

Kaufman, J. C., Gentile, C. A., & Baer, J. (2005). Do gifted student writers and creative writing experts rate creativity the same way? *Gifted Child Quarterly*, *49*, 260–265.

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. Hoboken, NJ: Wiley.

Kim, K., Lee, Y., & Lee, C. H. (2012). College students' style of language usage: Clues to creativity. *Perceptual and Motor Skills*, *114*, 43–50.

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., . . . Klemmer, S. R. (2015). Peer and self assessment in massive online classes. In H. Plattner, C. Meinel, & L. Leifer (Eds.), Design thinking research (pp. 131–168). New York, NY: Springer.

Lee, C. H., Kim, K. I., Lim, J. S., & Lee, Y. H. (2015). Psychological research using Linguistic Inquiry and Word Count (LIWC) and Korean Linguistic Inquiry and Word Count (KLIWC) language analysis methodologies. *Journal of Cognitive Science*, *16*, 132–149.

Lee, C. S., Huggins, A. C., & Therriault, D. J. (2014). A measure of creativity or intelligence? Examining internal and external structure validity evidence of the Remote Associates Test. *Psychology of Aesthetics, Creativity, and the Arts*, *8*, 446–460. https://doi.org/10.1037/a0036773

May, S. (2007). Doing creative writing. New York, NY: Routledge.

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*, 51–62.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*, 57–86.

McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, *23*, 35–59.

McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), Applied natural language processing: Identification, investigation, and resolution (pp. 188–205). Hershey, PA: Information Science Reference.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, *47*, 292–330. https://doi.org/10.1080/01638530902959943

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

Mozaffari, H. (2013). An analytical rubric for assessing creativity in creative writing. *Theory and Practice in Language Studies*, *3*, 2214–2219.

Mumford, M. D. (2003). Where have we been, where are we going? Taking stock in creativity research. *Creativity Research Journal*, *15*, 107–120.

Newman, J. (2007). The evaluation of creative writing at MA level (UK). In S. Earnshaw (Ed.), The handbook of creative writing (pp. 24–36). Edinburgh, UK: Edinburgh University Press.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*, 665–675.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, *43*, 121–152. https://doi.org/10.1080/01638530709336895

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, *9*, e115844. https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic Inquiry and Word Count: LIWC 2001. Mahway, NJ: Erlbaum.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2007). Linguistic Inquiry and Word Count: LIWC 2007. Austin, TX: LIWC.net.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment and Evaluation in Higher Education*, *35*, 435–448.

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, *15*, 18–39.

Rhodes, M. (1961). An analysis of creativity. *Phi Delta Kappan*, *42*, 305–310.

Robinson, R. L., Navea, R., & Ickes, W. (2013). Predicting final course performance from students' written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology*, *32*, 469–479.

Rodriguez, A. (2008). The "problem" of creative writing: Using grading rubrics based on narrative theory as solution. *International Journal for the Practice and Theory of Creative Writing*, *5*, 176–177.

Rogers, C. R. (1954). Toward a theory of creativity. *ETC: A Review of General Semantics*, *11*, 249–260.

Runco, M. A. (1988). Creativity research: Originality, utility, and integration. *Creativity Research Journal*, *1*, 1–7. https://doi.org/10.1080/10400418809534283

Silvia, P. J., Kaufman, J. C., & Pretz, J. E. (2009). Is creativity domain-specific? Latent class models of creative accomplishments and creative self-descriptions. *Psychology of Aesthetics, Creativity, and the Arts*, *3*, 139–148. https://doi.org/10.1037/a0014940

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., . . . Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*, 68–85. https://doi.org/10.1037/1931-3896.2.2.68

Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. *Handbook of Creativity*, *1*, 3–15.

Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, *105*, 409–429. https://doi.org/10.1037/0033-2909.105.3.409

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24–54.

Torrance, E. P. (2008). Torrance Tests of Creative Thinking: Norms-technical manual, verbal forms: A and B. Bensenville, IL: Scholastic Testing Service.

VanVoorhis, C. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, *3*, 43–50.

Varner, L. K., Roscoe, R. D., & McNamara, D. S. (2013). Evaluative misalignment of 10th-grade student and teacher criteria for essay quality: An automated textual analysis. *Journal of Writing Research*, *5*, 35–59.

Wolfradt, U., & Pretz, J. E. (2001). Individual differences in creativity: Personality, story writing, and hobbies. *European Journal of Personality*, *15*, 297–310.

Young, L. P. (2009). Imagine creating rubrics that develop creativity. *English Journal*, *99*, 74–79.

Zedelius, C. M., & Schooler, J. W. (2015). Mind wandering "Ahas" versus mindful reasoning: Alternative routes to creative solutions. *Frontiers in Psychology*, *6*, 834. https://doi.org/10.3389/fpsyg.2015.00834