Identifying Creativity During Problem Solving Using Linguistic Features

Stephen Skalicky, Scott A. Crossley, Danielle S. McNamara & Kasia Muldner

To cite this article: Stephen Skalicky, Scott A. Crossley, Danielle S. McNamara & Kasia Muldner (2017) Identifying Creativity During Problem Solving Using Linguistic Features, Creativity Research Journal, 29:4, 343-353, DOI: 10.1080/10400419.2017.1376490

To link to this article: http://dx.doi.org/10.1080/10400419.2017.1376490

Published online: 30 Oct 2017.
Identifying Creativity During Problem Solving Using Linguistic Features

Stephen Skalicky and Scott A. Crossley
Georgia State University

Danielle S. McNamara
Arizona State University

Kasia Muldner
Carleton University

Creativity is commonly assessed using divergent thinking tasks, which measure the fluency, flexibility, originality, and elaboration of participant output on a variety of different tasks. This study assesses the degree to which creativity can be identified based on linguistic features of participants’ language while completing collaborative divergent thinking tasks. To this end, 78 participants’ conversational dialogs (i.e., 39 dyads) within a chat environment were collected while completing three open-ended problem-solving tasks. Expert raters scored the dialogs in terms of fluency, flexibility, elaboration, and originality, as well as three types of creative language (metaphor and simile, humor, and word play). Factor analyses indicated that these scores captured two main constructs (creativity and elaboration). The linguistic features of the participants’ language (captured computationally using natural language processing tools) accounted for significant amounts of variation in both the creativity ($R^2 = .640$) and elaboration ($R^2 = .550$) scores within linear mixed effect (LME) models. These results highlight specific linguistic features that can be used to explain large amounts of variance in constructs related to creativity.

A common method of assessing creativity is through divergent thinking tasks, wherein participants generate as many solutions as possible to a prompt (e.g., generate different possible uses for a brick). The solutions are then scored for fluency (total number of ideas), flexibility (number of different idea categories), originality (uniqueness of ideas), and elaboration (explanation of ideas). These four scoring measures serve to quantify the “standard definition” of creativity (Runco & Jaeger, 2012, p. 92), which outlines two necessary components of creativity: originality and effectiveness. A number of studies have explored the validity and reliability of these metrics (Kaufman, Plucker, & Baer, 2008; Plucker & Makel, 2010; Runco & Charles, 1993; Runco & Mraz, 1992) including investigations into whether the metric of fluency acts as a “contaminating” factor for other metrics such as originality (Plucker & Makel, 2010, p. 55; Runco & Albert, 1985) and whether ideational pools of ideas are more valid than counting single ideas (Runco & Mraz, 1992).

Additionally, because several divergent thinking tasks necessarily involve the production of language, it is important to investigate the degree to which the linguistic features of the language used during these tasks correlate with assessments of creativity. Indeed, research has demonstrated that the semantic features of language produced during divergent thinking tasks associate strongly with measures such as flexibility and originality (Acar & Runco, 2014; Beketayev & Runco, 2016; Dumas & Dunbar, 2014). Other studies have also reported links between linguistic features and creativity in the domains of interviews (Djikic, Oatley, & Peterson, 2006; Forgeard, 2008), sarcasm

The results presented in this paper have not been used in any other reports or publications.

Acknowledgments: We thank Win Burleson for his role as PI on DRL-1319645. This work was supported by the Institute of Education Sciences [IESR305A080589, R305G20018-02] and National Science Foundation [DRL-1319645].

Correspondence should be sent to Stephen Skalicky, Department of Applied Linguistics/ESL, Georgia State University, 25 Park Place STE 1500, Atlanta, GA, 30303, USA. E-mail: scskalicky@gmail.com

CREATIVITY RESEARCH JOURNAL, 29(4), 343–353, 2017
Copyright © Taylor & Francis Group, LLC
ISSN: 1040-0419 print/1532-6934 online
DOI: https://doi.org/10.1080/10400419.2017.1376490

Downloaded by [Georgia State University] at 10:30 30 October 2017
(Huang, Gino, & Galinsky, 2015), and essay writing (Crossley, Muldner, & McNamara, 2016; Skalicky, Berger, Crossley, & McNamara, 2016). However, relatively little is known about the linguistic features of creative language produced during collaborative naturalistic dialogue.

The goal of this study is to examine links between language and creativity in naturalistic dialogue. Researchers have long recognized creative language as an important and ubiquitous element of communication. Although some researchers view linguistic creativity as specific tropes or linguistic forms, such as metaphor, word play, and humor (Cook, 2000; Gerrig & Gibbs, 1988; Martin, 2007; Veale, 2012), others have defined creative language as alignment between speakers involving a large amount of repetition during casual conversation (Carter, 2016; Carter & McCarthy, 2004). As such, the presence of another speaker may influence the types of linguistic creativity (e.g., originality, figurative language, humor) produced by the other speaker. Thus, this study examines the linguistic features of language produced during a collaborative problem-solving task to better understand connections between linguistic features and traditional measures of creativity (i.e., fluency, flexibility, originality, and elaboration), as well as linguistic measures of creativity (e.g., humor, figurative language, and alignment between speakers), within naturalistic discourse.

**LANGUAGE AND CREATIVITY**

One approach for identifying and assessing linguistic creativity is to examine various linguistic forms that constitute creative language and their corresponding pragmatic functions. Specific language tropes, such as metaphor, simile, wordplay, and humor are thought to be creative because they prompt a reconsideration of the surface meaning of an utterance (Gerrig & Gibbs, 1988; Martin, 2007; Ritchie, 2004) or manipulate (or play with) the morphological or semantic properties of language (Bell, 2005; Cook, 2000; Veale, 2012). Analyses of spoken conversation have identified specific pragmatic functions of these forms of creative language. For example, ironic criticism and praise (i.e., sarcasm) has been shown to serve a face-saving pragmatic function, mitigating the negative response associated with criticism (Jorgensen, 1996) or serving to downplay the overall positivity of a compliment or negativity of criticism (Dews, Kaplan, & Winner, 1995). Humorous language has been widely studied, with agreement that humor can serve to strengthen relationships and provide physical and psychological benefits (Bell, 2011; Martin, 2007). Importantly, creative language, compared to noncreative language, can impart specific pragmatic and semantic meanings more efficiently, and as a result, regularly occurs in normal conversation (Gerrig & Gibbs, 1988; Gibbs, 2000; Veale, 2012).

Work within discourse analytic perspectives has demonstrated that creativity in language includes not just explicit figures of speech described previously (such as metaphor and irony), but is also marked by larger, discourse-level features, such as repetition of words and ideas, as well as alignment between speakers (Carter, 2016; Carter & McCarthy, 2004), both of which can be measured linguistically through analysis of transcripts between two or more speakers. For example, Carter and McCarthy (2004) analyzed examples from the CANCODE corpus of spoken English and found many instances of both verbatim and varied repetition of clauses and phrases between speakers during spontaneous casual conversation. This repetition, they argued, functions to cocreate a shared frame of reference during a conversation that served to nurture the relationship between the speakers. They concluded that conversations are linguistically creative because speakers have an infinite number of choices when communicating, and that each new stretch of discourse produced can be creative if it is unique. When this process is carried out by two or more people, they share in this process and demonstrate their willingness to attend to each other’s linguistic creativity, albeit unconsciously.

Despite the current understanding regarding the linguistic forms and pragmatic functions of creative language, linguists have made little connection between creative language (i.e., linguistic creativity) and being creative. In other words, linguists have tended to analyze creative language as phenomena separate from noncreative language, without considering whether creative language is associated with greater creativity in general (e.g., during a divergent thinking task). However, creativity researchers have found several links between language and creativity. At a general level, individuals perceived as more creative will tend to produce language containing linguistic features that differ from less creative individuals. For example, Yu, Peng, Peng, Zheng, and Liu (2016) studied the effect of occupation on creative language output through an analysis of online microblogs (i.e., short online postings similar to Twitter) written in Mandarin Chinese. Yu et al. (2016) profiled the semantics of words produced by scholars, writers, business leaders, entertainment celebrities, sports stars, and other microbloggers. Their findings reported that scholars and writers produced more creative language than celebrities and sports stars.

Other research has demonstrated connections between figurative language use and creativity. For instance, Huang et al. (2015) conducted a series of experimental studies where participants created, interpreted, or recalled instances of sarcastic language. Their results indicated that participants exposed to sarcasm performed better on subsequent tests of creativity (both divergent and convergent tasks) as compared to participants not exposed to sarcasm. Huang et al. (2015) argued that exposure to sarcasm increased creative capacity because sarcasm enhanced the ability to think abstractly, which is a marker of creative problem solving.
Linguistic features have also been used to predict idea generation and humor in student writing. Crossley, Muldner, and McNamara (2016) examined links between measures of creativity (e.g., fluency, flexibility, originality, and elaboration) and human ratings of essay quality, as well as links between linguistic features in writing samples and overall ratings of idea generation. They found that essays rated as containing a greater number of flexible, original, and elaborated ideas were judged to be of higher quality. In addition, they found that linguistic features related to multiword units, word difficulty, semantic overlap, and word repetition were strong predictors of overall idea generation. In a similar study, Skalicky et al. (2016) examined links between human ratings of humor and essay quality and the degree to which linguistic features predicted humor ratings. They reported a weak correlation between humor and essay quality ($r = .195$). However, they also found that linguistic features related to descriptiveness (i.e., more adjectival and adverb use), lower cohesion (i.e., less paragraph-to-paragraph similarity), and greater lexical sophistication (lower word frequency) accounted for 17.5% of the variance in humor judgments.

More direct links have been made between creativity and language through investigations of linguistic features of the language produced during divergent thinking tasks. For example, Acar and Runco (2014) employed semantic networks to measure the distance between associations of ideas produced during divergent thinking tasks. Their findings demonstrated that participants rated higher for creativity made more remote associations between words. Dumas and Dunbar (2014) also used semantic distance measures to assess ideas produced during a divergent thinking task, and found that semantic distance provided a measure of originality just as reliable as human scoring. In another study using semantic distance methods, Beketayev and Runco (2016) found that semantic networks could successfully and automatically assess flexibility scores from divergent thinking tests.

**METHOD**

Participants

Data for this study came from computer-mediated conversations between sets of two participants (dyads). A total of 78 participants were recruited from a variety of different science, technology, engineering, mathematics, and arts disciplines, ranging from computer science to elementary education. Participants included 46 women, 31 men, and one participant who declined to identify gender. Fifty-five of the participants were enrolled in a 4-year undergraduate program, 17 were enrolled in masters or PhD programs, and six participants were recent graduates or nonstudents. Fifty of the participants spoke English as their first language. In total, 39 different dyads completed the three tasks.

Materials

Three divergent-thinking tasks from prior creativity research (e.g., Maimon & Horowitz, 1999) were used in this study. Each task presented an open-ended problem that asked the participants to develop as many solutions as they could think of to the problem (i.e., a divergent thinking task). For example, one task described a very flexible rubber pipe that needed to be cut in a perfectly straight line. However, the knife used to cut the pipe would bend the flexible pipe material while cutting it, preventing a straight cut. Participants were asked to think of as many solutions that would allow for a straight cut without bending the pipe. A full description of the three tasks can be found in Appendix A.

Procedure

Each pair of participants completed the three tasks in a counterbalanced order. Each participant worked with only one other participant. The participants were located at opposite ends of the same room, separated by room dividers so they could not see one another. To solve each task, participants communicated with one another through a short message service (SMS) interface on a computer analogous to Skype chat. The SMS assigned anonymous identification numbers to participants and logged all conversations.

**Human ratings**

Two trained raters evaluated the log transcripts from the dyad interactions using an analytic rubric developed to measure creativity. The rubric included seven different
subscles: four designed to capture idea generation related to cognitive aspects of creativity (fluency, flexibility, originality, and elaboration) and three designed to capture style related to linguistic types of creativity (humor, metaphor and simile, and word play). The analytic rubric used for human ratings of creativity can be found in Appendix B.

Each subscale of the rubric was rated on a scale of 1–6, with raters informed that the distance between each value on the scale was equal. Rater calibration occurred through initial rating of a random sample of 10 dyad interactions. For these interactions, raters only evaluated one of the three tasks and then compared their results, discussing and adjudicating any disagreements. Afterward, the raters independently scored the remainder of the interactions in a separate random order, focusing on one task at a time. After scoring was complete, the distance between rater scores was calculated for each subscale of the rubric and for each task. For any score with a distance greater than two points, raters discussed their scores and were given the option to adjudicate their scores based on the discussion. Correlations and Kappas for the raters’ scores after adjudication are reported in Table 1. All inter-rater reliability scores were within an acceptable range (i.e., \( r < .700 \), Kappa > .600). After rater adjudication, the scores for each rater were averaged and assigned to each of the individual tasks.

**Linguistic analysis**

The linguistic features of the dyad interactions were analyzed using three text analysis tools capable of measuring a wide variety of linguistic features related to lexical sophistication (The Tool for the Automatic Analysis of Lexical Sophistication, TAALES; Kyle & Crossley, 2015), cohesion (The Tool for the Automatic Analysis of Text Cohesion, TAACO; Crossley, Kyle & McNamara, 2016), and rhetorical features of writing quality (The Writing Assessment Tool, WAT; McNamara, Crossley, & Roscoe, 2013). TAALES reports on over 150 indices related to word frequency, range, academic words, word information measures, and n-gram features. TAACO reports on over 150 classic and revised indices of cohesion that evaluate cohesion on local (e.g., between sentences) and global levels (i.e., between paragraphs). WAT includes indices that overlap with TAALES and TAACO indices, but it also includes indices designed to measure rhetorical style in an essay (e.g., amplification words, hedge words, and exemplification words). The three constructs of interest (lexical sophistication, cohesion, and rhetorical features) are discussed more in depth below.

**Lexical sophistication.** Measures of lexical sophistication are based on word information measurements including word concreteness, imagability, familiarity, and meaningfulness, as well as lexical features related to word frequency and range (i.e., the number of texts a word occurs). Word information measurements tap into psycholinguistic properties of words. For example, word familiarity is a measure of how familiar a person is with a particular word, with dog being more familiar than obelisk. Word frequency measurements are based on spoken or written corpora and represent the relative use of a word in a language. Higher word frequency represents commonly used words, whereas low word frequency represents less commonly used words. Therefore, a text or conversation employing lower word frequency is indicative of larger vocabularies and more lexical sophistication. Word frequency measurements can also take into account the frequency of larger combinations of two (bigrams), three (trigrams) or more words (n-grams).

Lexical sophistication is an important component of creativity and linguistic style because it provides information describing the expansiveness and overall development of a speaker’s vocabulary (Kyle & Crossley, 2015). In tests of divergent thinking, the ability to access more words and concepts should allow for more overall ideas to be generated (Kharkhurin, 2012), while also providing more options for linguistic elements of style such as metaphor and humor (Gerrig & Gibbs, 1988).

**Cohesion.** Textual cohesion is related to the overlap of lexical and semantic content within a text and connections between text segments. Cohesion can be measured at multiple levels, including sentence-to-sentence and paragraph-to-paragraph. Textual features such as repeated words, the use of specific connective devices, and semantic similarity can all indicate higher or lower levels of cohesion in a text. Cohesion is important to investigate in creativity research because greater cohesion may signal less flexibility among ideas. Because ideational flexibility is dependent on the overall number of ideas produced, a higher number of similar or related ideas may result in lower flexibility scores (Runco, 1985, 2013). Additionally, measurements of textual cohesion may help capture less perceptible features of linguistic creativity described by discourse analytic perspectives, such as repetition and speaker alignment (Carter, 2016; Carter & McCarthy, 2004).

**Rhetorical devices.** Rhetorical devices are intended to capture common phrases and strategies employed by speakers to direct the flow of communication. For example,
language related to exemplification is typically prefaced by phrases such as for example. Because participants in this study were asked to generate as many ideas as possible, it may follow that specific rhetorical devices are used to control the flow and topics of discourse while engaged in creative endeavors. Additionally, participants may have preferred certain tenses and other grammatical choices to maintain a conversation facilitative of idea generation. A final consideration is that groups of words can be classified based on semantic category membership, such as social words (e.g., family, friends) or positive and negative emotional words (e.g., love, hate). In general, the rhetorical devices reported by WAT can capture phrasal, semantic, and grammatical elements of speech related to discourse management. Such features are important to investigate because idea generation and style in social exchanges may be influenced by certain tenses, phrases, and semantic categories (Fraser, 1999; Gumperz, 1982; Schiffrin, 1987; Tannen, 1981).

**RESULTS**

Factor Analysis

An exploratory factor analysis was first conducted to determine if the seven subscales should be collapsed into larger variables or factors, and, if so, to provide weighted scores that could then be used as a dependent variables in a statistical model. Because human ratings for the word play subscale were not normally distributed (in general, instances of word-play were absent, resulting in most transcripts receiving scores of zero), word play was not included in the factor analysis. A Bartlett’s test of sphericity was statistically significant (p < .001), suggesting that correlations between the subscales were not attributable to chance. The Kaiser-Meyer-Olkin measure of sampling adequacy reported .73, indicating underlying structure(s) (Field, 2013). The scree plot and percentage of variance explained between factors suggested a two-component extraction as the most parsimonious model. The rotated component matrix using varimax rotation with Kaiser normalization further supported a two-factor solution. When interpreting the components, the first component was labeled creativity based on the subscales that loaded into it (fluency, flexibility, originality, and humor). The remaining subscales (elaboration, metaphor, and simile) loaded into the second component, which was labeled elaboration, based on the strength of correlations reported within the component. All items loaded onto their subscales with eigenvalues > .700, see Table 2 for details.

Linear Mixed Effects Analysis

The indices reported by TAALES, TAACO, and WAT that lacked normal distributions (typically caused when an index included a high number of zeroes and ones in the output) were discarded, as transformation would not solve the high incidence of zeroes. Correlations were then calculated for the remaining variables to determine whether there was a statistical (p < .05) and meaningful (at least a small effect size, absolute r > .1) relation between the selected indices and the dependent variable (the component[s] reported by the factor analysis). Indices that were highly collinear (absolute r ≥ .900) were flagged, and the index with the strongest correlation with component scores was retained; the other indices were removed. To avoid over fitting and to control for repeated testing of the same data, any index with a p value greater than .001 was removed. All linguistic indices were transformed into standardized z-scores.

The remaining indices were included as predictor variables in LME models to determine which linguistic indices were significant predictors of the creativity scores. Models were built using R (R Core Team, 2015) with the lme4 (Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015) packages. For each model, the linguistic indices were entered as fixed effects (predictor variables), the component scores from the factor analysis were entered as the dependent variable, and participants and task types were both included as random effects. Additionally, task order was entered as a fixed effect to determine if participants became more or less creative as they continued to work on the tasks. The c.squaredGLMM function from the MuMin package (Nakagawa & Schielzeth, 2013), was used to obtain a measure of effect size. MuMin computes two R² values of variance explained: a marginal R² for the variance explained by the fixed factors (in this case, the linguistic indices and task order), and a conditional R² for the variance explained by both fixed and random factors (i.e., the linguistic indices, task order, and random variance attributed to the participants).

**Creativity and Elaboration Component Scores**

A weighted score was calculated for each dyad interaction (one per task) for the creativity component and the

<table>
<thead>
<tr>
<th>Subscales</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexibility</td>
<td>0.910</td>
<td>0.000</td>
</tr>
<tr>
<td>Originality</td>
<td>0.886</td>
<td>0.000</td>
</tr>
<tr>
<td>Fluency</td>
<td>0.851</td>
<td>0.000</td>
</tr>
<tr>
<td>Humor</td>
<td>0.831</td>
<td>0.000</td>
</tr>
<tr>
<td>Elaboration</td>
<td>0.907</td>
<td>0.000</td>
</tr>
<tr>
<td>Metaphor &amp; Simile</td>
<td>0.779</td>
<td>0.000</td>
</tr>
</tbody>
</table>

---

1 LME analyses were used in order to account for the random variance associated with subjects and items (see Baayen, Davidson, & Bates, 2008).
elaboration component. To do so, each individual subscale value (i.e., the averaged scores from the human ratings) was multiplied by its respective subscale’s eigenvalue. The results were then summed, resulting in a single weighted component score for creativity and elaboration for each of the three tasks each dyad completed. These weighted scores were used as dependent variables in the LME models.

Creativity Component Analysis

Correlational Analysis

In preparation for the LME analysis of creativity, correlations were conducted between the selected linguistic indices and the creativity component score. Of the available variables that were normally distributed, 21 demonstrated significant correlations with the creativity component score (based on \( p \leq .001 \)) while not demonstrating significant multicollinearity with other variables (absolute \( r > .899 \)). The correlations for these variables are reported in Table 3.

LME Analysis

The 21 variables that showed significant \( (p \leq .001) \) and meaningful correlations with the creativity component score were entered into a linear mixed effects model as fixed effects. The model reported that five of the 21 variables were significant predictors of the creativity component score. Based on their coefficients, three indices were identified as positive predictors of the creativity component score (lexical diversity \( D \), present tense words, and word meaningfulness) and two indices were identified as negative predictors (content word frequency and number of private verbs). This model reported a marginal \( R^2 \) of .640 and a conditional \( R^2 \) of .749, indicating these five variables collectively explained 64\% of the variance in the creativity component scores. Task order and task type were not significant, suggesting that the order and type of tasks did not account for significant variation in the creativity component scores. Table 4 reports the intercept and significant fixed effects, along with their coefficients, standard error, \( r \) values, and \( p \) values.

Elaboration Component Analysis

Correlational Analysis

In preparation for the LME analysis of elaboration, correlations were conducted between the selected linguistic indices and the elaboration component score. Of the available variables that were normally distributed, 30 demonstrated significant correlations with the elaboration component score (based on \( p \leq .001 \)) while not demonstrating significant multicollinearity with other variables (absolute \( r > .899 \)). The correlations for these variables are reported in Table 5.

LME Analysis

The 30 variables that showed significant and meaningful correlations with the elaboration component score were entered into the LME model as fixed effects. This model reported that six indices were significant predictors of the elaboration score. Based on their coefficients, four indices were identified as positive predictors (number of connectives, number of private verbs, number of exemplification phrases, and causal cohesion), while the remaining two indices were identified as negative predictors (word familiarity [sentence average] and number of unique bigrams). This model reported a marginal \( R^2 \) of .550 and a conditional \( R^2 \) of .768, indicating these six indices collectively explained 55\% of the variance in the elaboration component scores. As was the case with the creativity analysis, task order and task type were not significant predictors, suggesting the order and type of tasks did not account for significant variation in the Elaboration component scores. Table 6 reports the intercept...
linguistic features play in the assessment and performance on divergent thinking tasks. However, linguistic features are important elements of creativity as demonstrated in a number of domains, including divergent thinking tasks (Acar & Runco, 2014; Beketayev & Runco, 2016; Dumas & Dunbar, 2014), interviews (Dijkic et al., 2006; Forgeard, 2008), sarcasm (Huang et al., 2015), and essay writing (Crossley, Muldner, & McNamara, 2016; Skalicky et al., 2016). This study adds to previous research examining links between language and creativity during collaborative discourse using both traditional and linguistic definitions of creativity.

Creativity Component Scores

The factor analysis grouped fluency, flexibility, originality, and humor into the creativity component, suggesting that creativity can be explained using not only traditional measures of creativity (i.e., fluency, flexibility, and originality), but also by using humorous language. This is in line with previous humor research demonstrating that producing or being exposed to humorous language can increase one’s creativity (Martin, 2007).

In terms of the linguistic features predictive of the creativity component score, four linguistic features reported positive coefficients. Conversational dialogs receiving greater creativity component scores contained a wider range of vocabulary use (lexical diversity d), used higher amounts of present tense language (present tense words), and also used words with high numbers of associations (word meaningfulness). The two linguistic features with negative coefficients suggest that dialogs receiving higher creativity component scores included language with lower word frequency (content word frequency) and fewer verbs associated with internal thoughts and processes (number of private verbs).

More lexical diversity and lower word frequency are two direct markers of lexical sophistication (Kyle & Crossley, 2015). This suggests higher creativity component scores are associated with language that is more sophisticated, primarily because of a more developed vocabulary. Theoretically, these features relate directly to the constructs of ideational fluency (i.e., total number of ideas) and flexibility (i.e., total number of idea categories). If participants have access to a richer and more developed vocabulary, they can produce more and varied ideas. Indeed, Kharkhurin (2009, 2010) found that bilinguals performed differently compared to monolinguals on tests of divergent and convergent thinking because bilinguals have more words and concepts to draw from based on their larger linguistic repertoire (Kharkhurin, 2012). Therefore, ideational fluency and flexibility are potentially related to participants’ linguistic repertoire, which is reflected in their linguistic output.

Additionally, words with more associations (i.e., word meaningfulness) were also predictive of higher creativity component scores. To illustrate, the word dog is highly

and significant fixed effects, along with their coefficients, standard error, t values, and p values.

**DISCUSSION**

Divergent thinking tests are a common form of assessing creativity (Kaufman et al., 2008; Runco, 2013), but researchers do not often consider the potential role that

### TABLE 5

<table>
<thead>
<tr>
<th>index name</th>
<th>r</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrase coordination</td>
<td>.544</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Log frequency of trigrams (spoken)</td>
<td>.527</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of prepositions</td>
<td>.525</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of past tense words</td>
<td>.520</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of different function words (lemmas)</td>
<td>.502</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of infinitives</td>
<td>.489</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Incidence of modals</td>
<td>.478</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of gerund or present participle verbs</td>
<td>.475</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of private verbs</td>
<td>.467</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of exemplification phrases</td>
<td>.447</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of present tense words</td>
<td>.435</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Incidence of demonstratives</td>
<td>.431</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of repeated content words</td>
<td>.419</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of adjective attributive clauses</td>
<td>.417</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Semantic diversity</td>
<td>−.416</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Phrases connected by “and”</td>
<td>.404</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Incidence of first person pronouns</td>
<td>.394</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of conjuncts</td>
<td>.378</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Word concreteness</td>
<td>.365</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Maximum number of words before main verb</td>
<td>.361</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Incidence of subordinating conditionals</td>
<td>.354</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Word familiarity</td>
<td>−.349</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of amplifiers</td>
<td>.340</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of basic connectives</td>
<td>.300</td>
<td>.001</td>
</tr>
<tr>
<td>Casual cohesion</td>
<td>.299</td>
<td>.001</td>
</tr>
<tr>
<td>Incidence of not negations</td>
<td>.298</td>
<td>.001</td>
</tr>
<tr>
<td>Number of unique bigrams</td>
<td>−.294</td>
<td>.001</td>
</tr>
<tr>
<td>Incidence of logical connectives</td>
<td>.292</td>
<td>.001</td>
</tr>
<tr>
<td>Word imageability</td>
<td>.292</td>
<td>.001</td>
</tr>
<tr>
<td>Incidence of place adverbs</td>
<td>.291</td>
<td>.001</td>
</tr>
</tbody>
</table>

### TABLE 6

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>5.645</td>
<td>0.143</td>
<td>39.27</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Phrase coordination</td>
<td>0.538</td>
<td>0.102</td>
<td>5.263</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of private verbs</td>
<td>0.318</td>
<td>0.103</td>
<td>3.066</td>
<td>.002</td>
</tr>
<tr>
<td>Number of exemplification phrases</td>
<td>0.395</td>
<td>0.109</td>
<td>3.630</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Word familiarity (sentence average)</td>
<td>−0.323</td>
<td>0.102</td>
<td>−3.165</td>
<td>.002</td>
</tr>
<tr>
<td>Causal cohesion</td>
<td>0.279</td>
<td>0.108</td>
<td>2.593</td>
<td>.010</td>
</tr>
<tr>
<td>Number of unique bigrams</td>
<td>−0.394</td>
<td>0.107</td>
<td>−3.662</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
meaningful because it has many word associates (i.e., cat, pet, leash, bark, paws) whereas the word abscond has fewer word associates. Similar to the points made previously, higher word meaningfulness may facilitate idea generation because words with more associate meanings should allow for the generation of a greater number of ideas. Previous studies examining semantic network models of creativity have found similar results, in that semantic distance and association are predictive of divergent thinking scores (Acar & Runco, 2014; Beketayev & Runco, 2016; Dumas & Dunbar, 2014; Yu et al., 2016).

The tendency for transcripts with higher creativity component scores to remain in the present tense and avoid private verbs (e.g., think, know) is likely indicative of the nature of the task and in particular participants’ need to describe solutions to problems requiring physical manipulation (i.e., modifying an antenna, redesigning a baby chair, and cutting through a flexible pipe). To solve the problems, participants also needed to discuss hypothetical examples and, therefore, be less likely to recall events from the past or use private verbs that are less collaborative. Thus, participants who avoided private verbs and remained in present tense are likely those that stayed on task, collaborated better, and continued discussing solutions to the tasks.

**Elaboration Component Scores**

The elaboration component score derived from the factor analysis contained one subscale from traditional creativity measurements (i.e., elaboration) and one measure of linguistic creativity (i.e., metaphor and simile). Elaborating on ideas and metaphors have some commonalities. Elaboration is the extension of an idea (Kaufman et al., 2008) and metaphor is “understanding one kind of thing in terms of another” (Lakoff & Johnson, 1980, p. 5, emphasis in original). Thus, metaphor can serve to provide elaboration using figurative language, explaining why these two subscales loaded together into a single component.

As for the linguistic features predictive of the elaboration component score, four of the linguistic features were significant, positive predictors, suggesting transcripts with higher elaboration component scores contained more connections between independent clauses (phrase coordination), more private verbs, more phrases associated with exemplification, such as for example, and greater causal cohesion, which measures sentences connected by goal-oriented connective connectives such as and, but, and because and verbs such as cause, create, and make (Graesser, McNamara, & Kulikowich, 2011). The two indices with negative coefficients suggests transcripts with higher elaboration scores used language with lower word familiarity (i.e., how familiar and recognizable a word is) and lower bigram frequency (i.e., more infrequent two-word combinations), suggesting some measures of lexical sophistication also associated with higher elaboration scores.

In general, greater phrasal and causal coordination signals that conversational dialogs with higher elaboration component scores contained more ideas linked through words such as and, or, and because, signaling greater cohesion in the texts. This aligns well with reports from discourse analytic perspectives arguing repetition and speaker alignment are important elements of creative language use (Carter, 2016; Carter & McCarthy, 2004) because these indices suggest the same ideas were expanded upon by participants through clausal and phrasal coordination. Because elaboration is the extension of an idea, it follows that high amounts of coordination are necessary to elaborate on a certain idea or topic.

In contrast to what was found for the creativity component score, more private verbs were used during elaboration. This may be reflective of a rhetorical style of elaborating on ideas using conventions such as I think that, I wonder if, I remember that. This also suggests that elaboration can include discussion and verification of internal ideas and may be a feature of collaborative tasks in which participants discuss each other’s ideas. Similarly, higher elaboration scores were also marked by language related to providing examples, as evidenced by the greater presence of rhetorical devices related to exemplification (e.g., the phrase for example). This finding intuitively matches with what one would expect with elaboration on a topic. Providing examples about a topic not only helps clarify and explain it, but also keeps the conversation centered on the same topic.

Finally, lower word familiarity and bigram frequency suggests that, much like language associated with higher creativity component scores, elaboration includes an element of lexical sophistication. Specifically, higher elaboration component scores were associated with lower word familiarity and less frequent two-word combinations (bigrams). This may signal a need for elaboration to use more specific, less frequent language in order to elaborate on a particular topic or idea.

**Limitations**

A critical question is whether linguistic features merely help to identify creativity or whether they help to explain the construct itself? It is important to note that certain linguistic features account for variance in the rater scores, but not necessarily the creativity or elaboration of the participants who completed the divergent thinking task. In other words, the subscales of this rubric explain components of creativity (e.g., fluency, originality), and the linguistic features predict scores on those subscales. Nonetheless, consideration of the linguistic features selected by the LME models may provide insight into the linguistic features of the language associated with being creative or elaborative. For the creativity component scores, the findings indicate that vocabulary knowledge is linked to performance on divergent thinking tasks, and thus measures of creativity may be implicitly assessing vocabulary knowledge (and vice versa). To the same extent, vocabulary knowledge may also be an important element of
elaboration, as elaboration naturally requires being able to
discuss the same idea using different words. Indeed, because
other research has reported differences in vocabulary and
semantic use among different creative occupations (Acar &
Runco, 2014; Beketayev & Runco, 2016; Djikic et al., 2006;
Dumas & Dunbar, 2014; Yu et al., 2016), future research
using divergent thinking tasks may want to consider partici-
pants’ linguistic inventory using tests of vocabulary or reading
levels, and also consider the linguistic background of
participants (Kharkhurin, 2012), as participants who speak
more than one language may perform better due to a wider
linguistic repertoire.

It is also important to emphasize that our study ana-
lyzed data from collaborative tasks and that some of the
linguistic features predictive of the creativity or elabora-
tion component scores may be influenced by the interac-
tive nature of the task. Much of the divergent thinking
research is based on individual responses to divergent
thinking tests, whereas our findings may only apply to
collaborative contexts. Therefore, additional models
assessing individual and collaborative performance on
divergent thinking tests using a wider variety of linguistic
features should be developed to better determine which
features are the most generalizable to all divergent tests,
and which may be specific to individual and collaborative
divergent tests.

CONCLUSION

These results indicate that specific linguistic features
account for relatively high amounts of variance in rater
scores for both creativity (64%) and elaboration (55%) of
language produced during a collaborative divergent thinking
task. In addition, the results suggest that features representa-
tive of linguistic definitions of creativity, such as humor and
metaphor, are related to traditional metrics, such as
fluency and flexibility. As a whole, these results provide
creativity researchers with additional linguistic features to
explore when assessing performance on divergent thinking
tasks, while also further strengthening already known links
between language and creativity, especially for collaborative
tasks.

REFERENCES

ideas elicited by tests of divergent thinking. Creativity Research Journal,
26(2), 229–238. doi:10.1080/10400419.2014.901095
ing with crossed random effects for subjects and items. Journal of
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear
Retrieved from https://CRAN.R-project.org/package=lmme4
Beketayev, K., & Runco, M. A. (2016). Scoring divergent thinking tests by
computer with a semantics-based algorithm. Europe's Journal of
Psychology, 12(2), 210. doi:10.5964/ejop.v12i2.1127
study of humour in NS–NNS interaction. Applied Linguistics, 26(2),
192–218. doi:10.1093/applin/amh043
Bell, N. D. (2011). Humor scholarship and TESOL: Applying findings and
establishing a research agenda. TESOL Quarterly, 45(1), 134–159.
Carter, R. (2016). Language and creativity: The art of common talk (2nd ed.).
doi:10.1093/applin/25.1.62
Oxford University Press.
Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). Tool for the auto-
matic analysis of text cohesion (TAACO): Automatic assessment of
local, global, and text cohesion. Behavior Research Methods, 48(4),
1227–1237.
in student writing: Computational assessments and links to successful
0741088316650178
doi:10.1080/016385950944922
Djikic, M., Oatley, K., & Peterson, J. B. (2006). The bitter-sweet taste of
emoting: The linguistic comparison of writers and physicists.
s15326934crj1802_5
Dumas, D., & Dunbar, K. N. (2014). Understanding fluency and originality:
doi:10.1016/j.tsc.2014.09.003
Field, A. (2013). Discovering statistics using IBM SPSS Statistics: And
unipolar and bipolar mood disorder. Creativity Research Journal, 20(1),
81–92. doi:10.1080/10400410701842094
Fraser, B. (1999). What are discourse markers? Discourse Processes,
31(7), 931–952. doi:10.1080/016385959000101-5
Gerrig, R. J., & Gibbs, R. W., Jr. (1988). Beyond the lexicon: Creativity in
doi:10.1207/s15327876ms0301_1
Providing multilevel analyses of text characteristics. Educational
University Press.
Huang, L., Gino, F., & Galinsky, A. D. (2015). The highest form of
intelligence: Sarcasm increases creativity for both expressers and recipi-
ents. Organizational Behavior and Human Decision Processes, 131,
162–177. doi:10.1016/j.obhdp.2015.07.001
Pragmatics, 26(5), 613–634. doi:10.1016/0378-2166(95)
00067-4
Pipe Task. A very flexible rubber pipe has to be cut accurately. The cut must be across a straight line that is perpendicular to the surface of the rubber pipe. The current technology is to cut the pipe with a very sharp knife. The problem is that since the knife distorts the pipe before it starts penetrating it, the cut is not as accurate as required.

Collaborate with your partner to come up with as many solutions as you can think of to this problem. There are many, so don’t stop at one and remember, be creative!

Pipe Task. A very flexible rubber pipe has to be cut accurately. The cut must be across a straight line that is perpendicular to the surface of the rubber pipe. The current technology is to cut the pipe with a very sharp knife. The problem is that since the knife distorts the pipe before it starts penetrating it, the cut is not as accurate as required.

Work with your partner to come up with as many solutions as you can. Be creative!

Chair Task. Young children eat at a table by sitting in a baby chair, like the one shown on the right. Since tables


APPENDIX A. PROBLEM SOLVING TASKS

Antenna Task. A company has won a bid to design and manufacture a mobile military antenna that is to be handled and operated by a single soldier. The antenna system is composed of the antenna itself and a mast that supports it in a high position. The whole system is intended to be left for a period of time in one place, until it should be transferred to another.

Once the design was completed, it turned out that when ice accumulated on the antenna it became heavier, causing the mast to collapse (recall that no one is there to remove the ice). The obvious solution, to strengthen the mast through increasing its diameter, would result in a system that is too heavy for one soldier to carry. In summary, if the mast is strengthened it becomes too heavy, if it is not strengthened the antenna is bound to collapse.

Collaborate with your partner to come up with as many solutions as you can think of to this problem. There are many, so don’t stop at one and remember, be creative!

Pipe Task. A very flexible rubber pipe has to be cut accurately. The cut must be across a straight line that is perpendicular to the surface of the rubber pipe. The current technology is to cut the pipe with a very sharp knife. The problem is that since the knife distorts the pipe before it starts penetrating it, the cut is not as accurate as required.

Work with your partner to come up with as many solutions as you can. Be creative!

Chair Task. Young children eat at a table by sitting in a baby chair, like the one shown on the right. Since tables
have different heights, the gap between the height of a baby chair and the table might be too small or large causing inconvenience to the baby. How can you redesign the chair to fix this issue?

Work with your partner to come up with creative solutions to this problem. Discuss, collaborate, be creative!

Talk with your partner about various solutions – be creative!

APPENDIX B. CREATIVITY RUBRIC

Read each essay carefully and then assign a score on each of the points below. For the following evaluations, you will need to use a grading scale between 1 (minimum) and 6 (maximum).

We present here a description of the score as a guide using the example of does not meet the set criterion in any way versus meets the set criterion in every way. For example, a grade of 1 would relate to not meeting the criterion in any way, and a grade of 4 would relate to somewhat meeting the criterion. The distance between each grade (e.g., 1–2, 3–4, 4–5) should be considered equal. Thus, a grade of 5 (meets the criterion) is as far above a grade of 4 (somewhat meets the criterion) as a grade of 2 (does not meet the criterion) is above a grade of 1 (does not meet the criterion in any way).

<table>
<thead>
<tr>
<th>Score</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Does not meet the criterion in any way</td>
</tr>
<tr>
<td>2</td>
<td>Does not meet the criterion</td>
</tr>
<tr>
<td>3</td>
<td>Almost meets the criterion but not quite</td>
</tr>
<tr>
<td>4</td>
<td>Meets the criterion but only just</td>
</tr>
<tr>
<td>5</td>
<td>Meets the criterion</td>
</tr>
<tr>
<td>6</td>
<td>Meets the criterion in every way</td>
</tr>
</tbody>
</table>

Part

1. Ideas

1.1 Fluency
The essay contains many unique ideas within the essay.

1.2 Flexibility
The essay contains a variety of different ideas (e.g., many different categories of ideas).

1.3 Originality
The essay contains ideas that are unique across essays.

1.4 Elaboration
The essay includes information that expands on the main idea(s) contained in the essay.

2. Style

2.1 Humor
The essay attempts to provoke laughter or amusement.

2.2 Metaphor & Simile (cognitive style)
The essay involves original comparisons that construe entities outside of their content domain(s).

2.3 Word Play (linguistic style)
The essay includes the use of sounds, meanings, or forms of words that are unexpected or original.