Measuring Creative Ability in Spoken Bilingual Text: The Role of Language Proficiency and Linguistic Features

Stephen Skalicky (stephen.skalicky@vuw.ac.nz)
School of Linguistics and Applied Language Studies, Victoria University of Wellington
Wellington, New Zealand

Scott A. Crossley (scrossley@gsu.edu)
Department of Applied Linguistics and ESL, Georgia State University
Atlanta, Georgia, USA

Danielle S. McNamara (dsmcnama@asu.edu)
Department of Psychology, Arizona State University
Phoenix, Arizona, USA

Kasia Muldner (kasia.muldner@carleton.ca)
Institute of Cognitive Science, Carleton University
Ottawa, Ontario, Canada

Abstract

Whereas first language (L1) research has demonstrated that perceptions of creative ability are influenced by the complexity and diversity of language used to answer verbal tests of creativity, relatively little is known about the linguistic components of bilingual creative task performance. In this study, we analyze written transcripts of speech produced by 466 Japanese learners of English produced during a creative narrative task for features related to linguistic and cognitive dimensions of creativity. Then, we extract various linguistic features and test whether these features can predict human perceptions of creativity for the transcripts. Unlike L1 data, results suggest text length and L2 proficiency comprise the most parsimonious explanation of creativity scores in this L2 data. At the same time, linguistic features related to positive sentiment explained a significant yet small amount of additional variance in perceptions of creativity, suggesting texts with more positive language were perceived to be more creative.

Keywords: creativity, NLP, language proficiency, bilingualism

Introduction

The relationship between bilingualism and creativity can be approached from a number of perspectives. One is to investigate how learning a second language (L2) impacts creativity. Here, research has shown benefits of language learning, with high-proficiency bilinguals outperforming their monolingual and lower L2 proficiency peers on tests of creative ability (Kharkhurin, 2009; Leikin, 2013; Ricciardelli, 1992). Reasons for this difference have been attributed to the growth of language knowledge that naturally comes with mastering additional languages, suggesting that a specific cognitive ability (i.e., creativity) may be directly associated with language knowledge. Another approach to is to investigate the role of creativity in second language acquisition (SLA). For instance, researchers SLA have highlighted the facilitative role that creativity, play, and humor in an L2 can have on language learning (Cook, 2000; Pomerantz & Bell, 2007).

Yet another approach involves investigating links among creative ability, language use, and language knowledge in order to shed light on how language and cognition (specifically, creative ability) influence one another. One method for doing so, and the one that we adopt in the present work, is by determining whether linguistic features pattern with creativity.

The overarching objective of this study is to better understand how L2 proficiency and linguistic features relate to perceptions of creativity. To do so, this study examines linguistic features in 466 transcribed speech samples produced during an English L2 oral proficiency exam. The speech samples were part of the oral proficiency interviews found in the NICT Japanese Leaners of English (JLE) corpus (Izumi, Uchimoto, & Isahara, 2004; Tono et al., 2001). We trained raters to make creativity judgements for each of the samples. The linguistic features of the samples were then analyzed using automatic text analysis tools and associations between these features and the human judgments of creativity were assessed. This approach allowed us to examine the strength of the relations among L2 language proficiency, linguistic features of L2 speech, and expert raters' perceptions of creativity.

Creativity

Psychologists have defined creativity as a cognitive construct that represents the ability to develop novel and effective solutions to a problem (Kaufman, Plucker, & Baer, 2008; Runco & Jaeger, 2012). One common method for assessing creativity is through the use of divergent thinking tests, where a participant or group of participants generates as many

solutions to a problem that they can in a set amount of time (Runco, 2013). These tests are then most commonly scored for four measures: fluency (total number of ideas), flexibility (range of idea types), elaboration (ability to expand on ideas), and originality (uniqueness of ideas when compared to other participants' answers). In general, participants who score higher on these four features are thought to be more creative than those who score lower. Due to the frequent use of divergent thinking tests in creativity research, these four components have gained widespread acceptance as valid measures of creativity (Kaufman et al., 2008).

Bilingualism and Creativity

One specific application of divergent thinking tests has been to investigate whether bilinguals are more or less creative than monolinguals (Kharkhurin, 2009). A consistent finding from these studies is that the degree of bilingualism or relative proficiency in bilinguals' L2s is strongly related to creative performance (Kharkhurin, 2008). Specifically, language users with more balanced bilingualism (i.e., relatively similar proficiency between a user's two languages) significantly outperform those who report lower L2 proficiency compared to their L1 (Kharkhurin, 2011; Lee & Kim, 2011). These results have been replicated among different language users, including German-English and Dutch-English bilinguals (Hommel, Colzato, Fischer, & Christoffels, 2011) as well as among Hebrew-Russian bilingual children (Leikin, 2013).

Bilingual creative performance has also been identified as an important component of L2 learning. Specifically, language learners who experiment with the sounds, meanings, and forms of a language are a) better equipped to deduce the rules of a language, b) gain more agency over the language they are learning, c) construct more engaging learning environments, and d) enhance interaction with other learners (Bell, 2005; Cook, 2000). Although relatively high L2 proficiency is required to take part in complex forms of language play such as interpersonal humor (Bell, 2005), even lower proficiency L2 learners have demonstrated usage of less complex forms of play (Bell, Skalicky, & Salsbury, 2014).

Linguistic Features, Bilingualism, and Creativity

Learning a second language naturally involves increased knowledge of lexical items and word associations in that language. In English, lexical features such as polysemous word senses, hypernymic categories, and psycholinguistic measures of lexical sophistication have all been shown to change over time as learners increase their L2 English proficiency. Specifically, as L2 English learners become more proficient, they develop more polysemous and less frequent senses for English words (Crossley, Salsbury, & McNamara, 2010), more diverse hypernymic relations among word categories in English (Crossley, Salsbury, & McNamara, 2009), and demonstrate higher levels of lexical sophistication in English (e.g., more abstract lexical items that are less rooted in the immediate context; Salsbury et al., 2011).

Several of these same linguistic features have been associated with higher performance on tests of creativity in English as an L1. For example, words generated by individuals rated higher for creativity have more remote associations among concepts as measured through computationally-derived association strengths such as Latent Semantic Analysis (Acar & Runco, 2014; Beketayev & Runco, 2016; Dumas & Dunbar, 2014). Higher creativity scores are also associated with higher levels of lexical sophistication (i.e., more infrequent, varied, and complex language) and semantic cohesion (Skalicky, Crossley, McNamara, & Muldner, 2017).

Current Study

The current study has two goals. The first is to examine the extent to which L2 English proficiency is associated with perceptions of creativity during an oral picture description task among Japanese-L1 English-L2 bilinguals of eight different L2 proficiency levels. The second is to investigate whether linguistic features of the language produced during the task are predictive of perceptions of creativity. Because SLA research has demonstrated that various aspects of language such as lexical sophistication change over time as one gains proficiency in English as an L2, we examine the extent that differences in creative output based on L2 proficiency are associated with quantifiable features of language. By identifying language features associated with perceptions of creativity, we aim to further define linguistic aspects of L2 creativity, identify associations between creativity and proficiency in a second language, and provide additional explanations for differences in creative performance among bilinguals of differing proficiency levels. The following research questions guide our study:

1. What role does L2 English language proficiency have for human perceptions of creativity during an English L2 oral proficiency exam?
2. Do linguistic features explain differences in creativity scores when controlling for L2 English proficiency?

Method

Corpus

We used a subset of the NICT Japanese Learner English Corpus to collect creativity ratings for L2 speakers of English (Izumi et al., 2004). The JLE comprises over 1200 recorded speech samples of Japanese learners of English who completed an interview activity designed to assess their oral English proficiency. The JLE data also includes the oral proficiency scores for each interviewee assigned by the interviewer at the time of the interview. The scores were derived using the Standard Speaking Test scoring method (Tono et al., 2001), where 2-3 raters used a holistic rubric based on the American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines to place

interviewees into one of nine different levels based on their oral proficiency (1 being the lowest and 9 being the highest).

Each interview was conducted between a test taker and a test administrator. The interviews lasted approximately 10-15 minutes and included three interview tasks. In this study, we focus on the final interview task, which was a picture description task where the interviewee was asked to construct a story based on information depicted in a picture or a set of pictures. We focused on this task because it provided the strongest potential for the test takers to produce creative ideas in that they were given the freedom to embellish and elaborate on events in the story as they constructed it. The interviewer provided minimal feedback beyond confirmation checks and backchanneling, ensuring that all the ideas produced during this task belonged to the interviewee. Within the picture sequence description task, there were ten possible picture sets that depicted scenes such as camping, visiting a zoo, eating at a restaurant, and shopping in a grocery store.

We constructed a subset of the JLE corpus by randomly selecting 250 texts from male and female speakers respectively ($N = 500$) while also sampling equally from each proficiency level (levels 2-9 with the exception of 1, which was rare). For each file we manually removed all text not associated with the picture sequence description task and all speech produced by the interviewer, leaving just the text that was on topic and delivered by the test taker. In order to ensure enough coverage for our linguistic measurements, we further removed any text containing less than 50 words (34 texts), resulting in a final JLE subset of 466 texts (237 female, 229 male). The average number of words per text in this final subset was 140.700 ($SD = 62.551$). The resulting distribution of proficiency levels approximated a normal distribution ($M = 5.361$, $SD = 1.744$).

Human Ratings

We developed an analytic rubric to obtain creativity ratings for each text in our dataset. The rubric contained seven different subscales with a range of 1 (does not meet the criterion in any way) to 6 (meets the criterion in every way). The subscales were divided into two larger categories: IDEAS and STYLE. The IDEAS category contained four subscales related to cognitive definitions of creativity: ideation (the speaker produced a large number of different ideas), originality (the speaker's ideas were original when compared to other speakers completing the same task), elaboration (the speaker included additional information elaborating on their ideas) and appropriateness (the speaker's ideas created an effective narrative). The STYLE category contained three subscales related to linguistic creativity: humor (the speaker produced at least one idea intending to provoke humor or amusement), metaphor and simile (the speaker produced ideas which made conceptual comparisons), and word play (the speaker played with the sounds or meanings of words).

Two native English-speaking research assistants were trained on the creativity rubric using a separate subset of 65 JLE texts. Raters were informed that the distance between each number on the rating scale was equal. After calibrating on the initial 65 texts, the raters then independently scored the remaining 466 texts for creativity. The raters were not aware that the samples were from English L2 learners. After scoring, raters were able to adjudicate disagreements greater than two for any of the subscales. Raters reported almost no instances of humor, metaphor and simile, or wordplay in the corpus, and thus these subscales were removed from the study. Table 1 displays the final, adjudicated kappa scores and correlations between the two raters for each of the remaining five subscales. After adjudication, the raters' scores were averaged for each subscale and text.

Table 1: Rater agreement

Subscale	*r*	Kappa
Ideational Fluency	0.830	0.830
Originality	0.825	0.822
Elaboration	0.739	0.738
Appropriateness	0.785	0.781

Linguistic Feature Selection

Based on prior work reporting associations between lexical sophistication, cohesion, and creativity in L1 English research (Acar & Runco, 2014; Beketayev & Runco, 2016; Dumas & Dunbar, 2014; Skalicky et al., 2017), we hand-selected a range of lexical indices representative of these constructs. We also included features related to sentiment in order to explore whether these measures might explain further explain variance in creativity scores. We obtained our measures of lexical sophistication, sentiment, and cohesion using three freely-available automatic text analysis tools, TAALES v2.2, SEANCE, and TAACO, respectively (see Crossley, Kyle, & McNamara, 2016a, 2016b; Kyle, Crossley, & Berger, 2017).

For lexical sophistication, we included linguistic indices of word frequency, word concreteness (i.e., how abstract a word's meaning is), contextual diversity and distinctiveness (i.e., the range of different contexts a word occurs in), word meaningfulness (i.e., number of associations with other words), word polysemy (i.e., the number of different senses a word form has), and word recognition and naming norms (i.e., average time to recognize and name English words). For cohesion, we included features measuring the type-token ratio (i.e., lexical diversity) and number of repeated content words in each text. Finally, for sentiment, we used features measuring the overall valence of a text (i.e., use of positive or negative vocabulary). We used measures calculated for content words (e.g., nouns, verbs, adjectives) only.

Statistical Analysis

We first conducted a principal component analysis using the raters' scores for the four subscales in the IDEAS category from the creativity rubric to develop a single, weighted creativity score to be used as the dependent variable. We then conducted correlations between the creativity score and the oral proficiency scores provided with the JLE corpus, as well as between the creativity score and text length (i.e., number

of content word types in each text). We included text length as a variable because longer texts would include more ideas and thus be biased to higher ideation scores (and therefore higher creativity scores). Then, we controlled the linguistic features based on correlations with the dependent variable and also controlled for multicollinearity using correlations and variance inflation factors.

Next, in order to test whether L2 proficiency and the linguistic features related to lexical sophistication, sentiment, and cohesion were predictive of the creativity scores, we performed comparisons between linear regression models in order to obtain the most parsimonious model (i.e., the model that explained the largest amount of variance with the fewest number of predictor variables).

Results

Principal Component Analysis

A principal component analysis (PCA) was conducted on the averaged ratings of ideation, originality, elaboration, and appropriateness from the analytic rubric for the 466 texts in our subset of the JLE corpus. A Bartlett's test of sphericity was statistically significant (χ^2 = 717.179, df = 6, p < .001), and the Kaiser-Meyer-Olkin measure of sampling adequacy reported .672, representing acceptable ability for the PCA to yield distinct, reliable factors (Field, 2013). A single component containing all four variables accounted for 59.463% of cumulative variance with an eigenvalue of 2.378. The individual subscale loadings were: ideation = .913, elaboration = .892, appropriateness = .782, originality = .369. In order to calculate a single score reflective of the different strengths of these loadings we multiplied each human score for each subscale for each text by its respective loading and summed these values per text, obtaining a weighted sum component score for each text (DiStefano, Zhu, & Mîndril, 2009), which we refer to as the creativity score (Min = 8.238, Max = 14.780, M = 12.526, SD = 1.545).

Linguistic Feature Reduction

Using the output from the automatic text analysis programs, we first reduced the number of variables by only including variables of interest that had a significant and meaningful linear relation (i.e., absolute r > .1) with the dependent variable (i.e., the creativity score). We then controlled for multicollinearity using variance inflation factors (VIF), removing any variable with a VIF greater than 2. The end result was a selection of seven linguistic indices that demonstrated no strong multicollinearity and possessed a significant linear relation with the dependent variable.

These features were: average Age of Acquisition, which is based on averaged self-reported ratings of the age English users first understood 30,000 different English words collected from over 800,000 English speakers in the United States (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), average Spoken Word Frequency calculated from the Corpus of Contemporary American English, The University of South Florida Free Association Norms (i.e., the average number of words a subject voices when presented with a particular word; Nelson, McEvoy, & Schreiber, 1998), LSA Average Top Three Cosine (average LSA cosine values for the top three related words in each text), Vader Positive Sentiment (compound score measuring the overall positive sentiment in a text; Hutto & Gilbert, 2014), Number of Content Word Types (our measure of text length), and Number of Repeated Content Word Lemmas (divided by total text length). Table 2 displays these variables and their correlations with the creativity score, along with the correlation between L2 proficiency level and creativity.

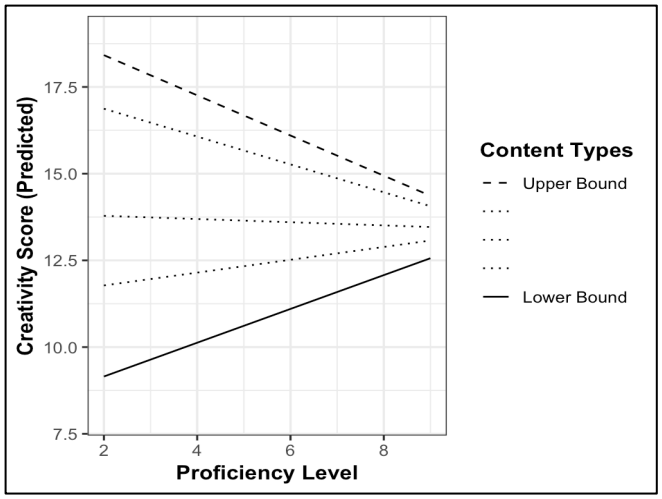

Figure 1: Interaction between English proficiency and text length (number of content word types). Upper and lower bounds represent the minimum and maximum values for number of content word types.

Regression Models

Based on the correlations among creativity, L2 proficiency level, and text length and initial model exploration, the results suggested a large amount of the variance in raters' creativity scores could be captured in a linear regression model fit with L2 proficiency, text length, and an interaction between text length and L2 proficiency. This model explained approximately 48% of the variance in raters' creativity scores (R^2 = .475, F[3, 462] = 141.100). The significant interaction between text length (i.e., number of content word types) and L2 proficiency indicated that differences in text length at higher L2 proficiency levels had significantly less effect on raters' perceptions of creativity when compared to lower levels of L2 proficiency. Specifically, at lower levels of L2 proficiency, texts with a higher number of content word types were rated significantly higher for creativity, and this effect attenuated significantly at higher levels of L2 proficiency. This interaction is visually plotted in Figure 1, and Table 3 displays the standardized beta coefficients and 95% confidence intervals for the terms in the model.

We then tested whether the separate inclusion of each of the remaining six predictor variables would significantly improve the baseline model based on changes in adjusted R^2 by comparing different linear regression models using the

anova() command in R. Table 4 summarizes the results of these comparisons. As can be seen, only the Vader Positive Sentiment index significantly increased the adjusted R^2 of the baseline model, with an increase of .07% variance explained. The remaining linguistic features did not explain any significant amount of additional variance, further suggesting that perceptions of creativity were strongly associated with L2 proficiency level and the amount of text produced by each participant.

Table 2. Correlations between predictor variables and the creativity score.

Index	*r*
Number of Content Word Types	0.643
L2 English Proficiency Level	0.449
Average Age of Acquisition	0.288
LSA (mean top three cosine)	-0.281
Free Association Norms (USF)	-0.240
Vader Positive Sentiment	0.158
Repeated Content Lemmas	0.134
Spoken Word Frequency (COCA)	0.118

Discussion

Creativity and L2 English Proficiency

Our first research question asked whether L2 English proficiency influenced raters' perceptions of creativity among our speech samples. The moderate correlation between the creativity score and L2 proficiency in Table 2 suggests a positive association between these features. This is further supported by the baseline regression model (Table 3), which included a significant positive effect for L2 proficiency (moderated by text length, see below). Together, these results provide an additional piece of evidence suggesting that a higher L2 proficiency level is associated with greater perceptions of creativity among the creativity raters. This finding aligns well with prior research into bilingual creative performance, which also reported greater creativity levels among bilinguals with higher L2 proficiency (Hommel et al., 2011; Kharkhurin, 2011; Lee & Kim, 2011; Leikin, 2013).

Our findings also suggest that this effect was moderated by text length, in that the overall length of the participants' picture description narratives (i.e., number of content word types) was more strongly associated with raters' perceptions of creativity at lower compared to higher levels of L2 proficiency. Thus, the manifestation of L2 proficiency as the ability to produce more language may be the driving determinant between higher creativity scores and L2 proficiency, as the ability to produce more language allowed for the opportunity to produce more ideas, and therefore receive higher ideation ratings and thus higher creativity scores.

Table 3: Baseline model explaining variance in raters' perceptions of creativity.

Model Term	*Estimate*	*SE*	*t*	*p*	*5% CI*	*95% CI*
(Intercept)	11.977	0.196	61.140	< .001	11.654	12.300
L2 English Proficiency Level	0.132	0.035	3.804	< .001	0.075	0.189
Text Length	1.900	0.176	10.788	< .001	1.609	2.190
L2 English Proficiency Level * Text Length	-0.178	0.028	-6.255	< .001	-0.225	-0.131
Adjusted R^2 = .475, F(3, 462) = 141.100. Estimate represents standardized beta coefficient as all predictor variables were z- scored before being entered into the model.						

Table 4: Comparisons between baseline model and models with different linguistic features.

Model Term	R^2	Adjusted R^2	*F*	*p*	R^2 Difference (Adjusted)
Baseline Model	0.478	0.475	141.099	NA	NA
Vader Positive Sentiment	0.487	0.482	109.275	0.006	0.007
Spoken Word Frequency (COCA)	0.478	0.474	105.597	0.962	0.001
Free Association Norms (USF)	0.480	0.476	106.513	0.167	0.001
LSA (mean top three cosine)	0.478	0.474	105.657	0.721	0.001
Age of Acquisition	0.478	0.474	105.677	0.680	0.001
Repeated Content Lemmas	0.478	0.474	105.684	0.667	0.001
Note: DF for all comparison models = (4, 461). Baseline model R syntax = *creativity ~ L2 English Proficiency Level + Text Length + L2 English Proficiency Level:Text Length*. *F* and *p* values correspond to change in R^2 from baseline model.					

Linguistic Features and English Proficiency

Our second research question asked whether linguistic features explained differences in creativity scores while taking L2 English proficiency into account. Early model exploration as well as a series of hierarchical linear regression comparisons suggested that almost all of the linguistic features selected for this study failed to predict any

meaningful amount of variance beyond the effect of L2 proficiency and text length, which combined to explain nearly 50% (Adjusted R^2 = .475) of the variance in the raters' perceptions of creativity, suggesting a relatively strong effect; see Table 4. The interaction between L2 proficiency and text length demonstrates that while text length was a strong, significant predictor of creativity scores, this effect was significantly stronger at lower L2 proficiency levels. Specifically, while texts with a greater number of content word types predicted increased creativity scores for participants across all eight proficiency levels, this effect was much stronger for participants who received lower L2 proficiency scores by the interviewer.

When comparing the difference in creativity scores between the upper bound and the lower bound of text length in Figure 1 (i.e., the minimum and maximum values for number of content word types), this difference attenuates for participants with higher L2 proficiency scores. This suggests that while differences in creativity scores at lower L2 proficiency levels are strongly predicted by the ability to produce more words (and therefore more ideas), this was not the case at the higher L2 proficiency levels. At higher L2 proficiency levels, variation in creativity scores based on total number of content word types was relatively low, suggesting that other features of the texts may have influenced the raters' creativity scores at higher L2 proficiency levels. However, these additional features, if any, were not captured in any of the linguistic features provided by our automatic text analysis tools. Therefore, unlike results reported in the L1 data, it is difficult at this time to draw concrete connections between specific linguistic features and bilingual performance on tests of creativity. It may be the case that additional linguistic features not included in the current study can explain variance in creativity at higher levels of L2 proficiency, providing ample opportunity for future research.

Aside from text length, one index, Vader Positive Sentiment, did result in a significantly better regression model fit, but only by approximately .07% of variance explained, suggesting that this index had a relatively weak effect. Nonetheless, it is still worth considering why this index may have provided a significant amount of additional variance explained. The Vader Positive Sentiment index is a component score derived from formulas specifically designed to measure sentiment in shorter texts, especially those used in social media (Hutto & Gilbert, 2014). The coefficient for the Vader Index was .148 (intercept = 11.881), suggesting a positive relation between positive sentiment and perceptions of creativity. Thus, narratives with more positive vocabulary may have appeared more creative to the raters in this study. Perhaps narratives with more positive language reflects a greater intent by the speakers in the corpus to create a unique story, as compared to narratives that were more factual descriptions of events. It would thus be worthwhile to further consider the role of sentiment in linguistic investigations of creative performance, as this would help identify links between specific types of linguistic knowledge and the cognitive construct of creativity.

Conclusion and Limitations

Previous investigations of bilingual creativity have reported a tendency for bilinguals with greater L2 proficiency to outperform those with lower L2 proficiency on standardized tests of creativity. The results from the current study support these claims while raising further questions. Specifically, we observed that increased levels of L2 proficiency were associated with higher perceptions of creativity, but this effect was moderated by the length of the speech samples. Moreover, while our results identified Vader Positive Sentiment as a significant linguistic predictor of creativity, this (and our other linguistic features) was overshadowed by the strong effect of text length. As a whole, these results suggest that there may be an L2 proficiency threshold for bilingual creativity, in that raters attended to additional linguistic features beyond text length only for speakers with relatively higher levels of L2 proficiency. In the future, it may be helpful to incorporate diversity-based linguistic information based on the prompts in order to control for potential vocabulary differences among the different prompts, which may influence the raters' perception of creativity (Chiru & Rebedea, 2017).

One final consideration is that the L2 English proficiency measure used in the current study was based solely on oral L2 proficiency at the time of the picture description task. Previous research in bilingual creativity has relied on proficiency assessments based on vocabulary knowledge tests as well as participant self-ratings of L2 proficiency and levels of bilingualism, which captures receptive vocabulary knowledge (i.e., reading and listening ability). The JLE L2 proficiency scores, on the other hand, are a measure of productive vocabulary knowledge (i.e., speaking and writing skill), and productive vocabulary size is typically smaller than receptive size (Schmitt, 2008). However, receptive and productive vocabulary knowledge are inextricably linked, suggesting that the JLE oral proficiency score is also a correlate of receptive L2 vocabulary knowledge (Webb, 2008). In all, these findings further highlight the association between bilingualism and the cognitive ability of creativity while providing avenues for future research.

References

Acar, S., & Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal*, *26*, 229–238.

Beketayev, K., & Runco, M. A. (2016). Scoring divergent thinking tests by computer with a semantics-based algorithm. *Europe's Journal of Psychology*, *12*, 210.

Bell, N. D. (2005). Exploring L2 language play as an aid to SLL: A case study of humour in NS-NNS interaction. *Applied Linguistics*, *26*, 192–218.

Bell, N. D., Skalicky, S., & Salsbury, T. (2014). Multicompetence in L2 language play: A longitudinal case study. *Language Learning*, *64*, 72–102.

Chiru, C.-G., & Rebedea, T. (2017). Profiling of participants in chat conversations using creativity-based heuristics. *Creativity Research Journal*, *29*, 43–55.

Cook, G. (2000). *Language play, language learning*. Oxford University Press.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016a). Sentiment analysis and social cognition engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*, *49*, 803–821.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016b). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, *48*, 1227–1237.

Crossley, S. A., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, *59*, 307–334.

Crossley, S. A., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, *60*, 573–605.

DiStefano, C., Zhu, M., & Mîndril, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation*, *14*, 1–11.

Dumas, D., & Dunbar, K. N. (2014). Understanding fluency and originality: A latent variable perspective. *Thinking Skills and Creativity*, *14*, 56–67.

Field, A. (2013). *Discovering statistics using IMB SPSS Statistics: And sex and drugs and rock "n" roll* (4th ed.). Sage.

Hommel, B., Colzato, L. S., Fischer, R., & Christoffels, I. K. (2011). Bilingualism and creativity: Benefits in convergent thinking come with losses in divergent thinking. *Frontiers in Psychology*, *2*, 1–5.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, 216–225.

Izumi, E., Uchimoto, K., & Isahara, H. (2004). The NICT JLE Corpus: Exploiting the language learner's speech database for research and education. *International Journal of the Computer, the Internet and Management*, *12*, 119–125.

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

Kharkhurin, A. V. (2008). The effect of linguistic proficiency, age of second language acquisition, and length of exposure to a new cultural environment on bilinguals' divergent thinking. *Bilingualism: Language and Cognition*, *11*, 225–243.

Kharkhurin, A. V. (2009). The role of bilingualism in creative performance on divergent thinking and invented alien creatures tests. *The Journal of Creative Behavior*, *43*, 59–71.

Kharkhurin, A. V. (2011). The role of selective attention in bilingual creativity. *Creativity Research Journal*, *23*, 239–254.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990.

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods*, *50*, 1030–1046.

Lee, H., & Kim, K. H. (2011). Can speaking more languages enhance your creativity? Relationship between bilingualism and creative potential among Korean American students with multicultural link. *Personality and Individual Differences*, *50*, 1186–1190.

Leikin, M. (2013). The effect of bilingualism on creativity: Developmental and educational perspectives. *International Journal of Bilingualism*, *17*, 431–447.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms [Database]*. Retrieved from w3.usf.edu/FreeAssociation.

Pomerantz, A., & Bell, N. D. (2007). Learning to play, playing to learn: FL learners as multicompetent language users. *Applied Linguistics*, *28*, 556–578.

Ricciardelli, L. A. (1992). Creativity and bilingualism. *The Journal of Creative Behavior*, *26*, 242–254.

Runco, M. A. (Ed.). (2013). *Divergent thinking and creative potential*. New York, NY: Hampton Press.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*, 92–96.

Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, *27*, 343–360.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363.

Skalicky, S., Crossley, S. A., McNamara, D. S., & Muldner, K. (2017). Identifying creativity during problem solving using linguistic features. *Creativity Research Journal*, *29*, 343–353.

Tono, Y., Kaneko, T., Isahara, H., Saiga, T., Izumi, E., Narita, M., & Kaneko, E. (2001). The Standard Speaking Test (SST) Corpus: A 1 million-word spoken corpus of Japanese learners of English and its implications for L2 lexicography. *Second Asialex International Congress, Korea*, 257–262.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, *30*, 79–95.