Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference

Automatically Identifying Humorous and Persuasive Language Produced during a Creative Problem-Solving Task

Stephen Skalicky,[a] Scott A. Crossley,[a] Danielle S. McNamara,[b] Kasia Muldner[c]

[a]Department of Applied Linguistics, Georgia State University

[b]Department of Psychology, Arizona State University

[c]Institute of Cognitive Sciences, Carleton University

Abstract

Despite being an important component of problem-solving ability, relatively little is known about the linguistic features of creativity and the related constructs of elaboration, humor, and persuasion. In order to better understand the linguistic features of these constructs, two analyses were performed to examine relationships between linguistic features and human judgments of humor, persuasion, creativity, and elaboration. First, linguistic indices derived from automatic text analysis tools were used to predict human categorizations of utterances for humor and persuasion in a corpus of natural dialogue produced during a creative problem-solving and divergent thinking task. Four linguistic indices related to the use of function words, word age-of-acquisition, and spoken word frequency distinguished humor and persuasion with approximately 50% accuracy. These linguistic features, along with incidence scores for humorous and persuasive categories based on human ratings per dialogue were then used to predict human ratings of creativity and elaboration from the same corpus. Less variation in use of function words significantly predicted both creativity and elaboration scores, and lower spoken word frequency and higher incidences of humorous utterances significantly predicted creativity scores. These results demonstrate the potential for linguistic features to explain creativity and elaboration and highlight connections between humor, persuasion, and creativity.

Introduction

Humor, persuasion, and creativity are important components of linguistic ability, problem solving, and writing proficiency (Bench-Capon, 2003; Crossley, Muldner and McNamara, 2016; Runco 2013). In this study, we use natural language processing (NLP) techniques to investigate relations between linguistic features and humorous and persuasive language produced by participants completing a collaborative problem-solving and divergent thinking task. We focus on identifying instances of humor and elaboration/persuasion because they are two conversational strategies that can affect the overall creativity of discourse (Bench-Capon 2003; Cundall 2007). We link humor and elaboration/persuasion to creativity in order to better understand the linguistic mechanisms that underlie creativity.

Creativity

Research into creativity has traditionally relied on assessment techniques that focus on the number and quality of ideas produced by participants during divergent thinking tasks (Kaufman, Plucker, and Baer 2008). Divergent thinking tasks ask participants, either individually or in pairs, to produce lists that contain as many creative solutions as possible to a problem (e.g., think of as many uses as possible for a brick; Runco 2013). In general, participants in divergent thinking tasks are rated as more creative if they produce more ideas, ideas that are more original, and ideas that are more effective for solving a task (Kaufmen et al. 2008). In addition, research shows that while the processes of solving divergent thinking tasks between individuals and pairs are different, there are no discernable differences between individuals and pairs in the creativity of their responses (Tidikis and Ash 2013). While this research has provided insight into the cognitive nature of creativity, there has been relatively little examination of the linguistic features associated with the creativity of answers elicited during divergent thinking tests (cf. Acar and Runco, 2014; Skalicky et al. 2016b). One reason for this is that the output of divergent thinking tasks is a list of solutions, but not the natural discourse that led to the list of solutions. The unanalyzed discourse underlying the list of solutions is a crucial component of linguistic analysis because it contains the natural language output that led to the solutions.

Thus, when two or more participants are asked to collaboratively complete a divergent thinking task together, analyzing the actual discourse that leads to creative solutions will introduce new interactional features of language that are not found in the lists of creative solutions that are generally analyzed in divergent thinking tasks. Analyzing the actual discourse will afford the opportunity to examine if certain linguistic strategies and features within the discourse pattern with higher or lower ratings of creativity. Links between linguistic features in discourse and creativity have been reported in previous research indicating that measures of lexical sophistication are predictive of creativity in problem-solving tasks. Specifically, language judged to be more creative in problem solving tasks contained higher lexical diversity (i.e., more varied vocabulary), lower word frequency (i.e., less frequently used words), and more word associations (i.e., words with more links to other words; Skalicky et al. 2016b).

Humor

Humor is a creative form of language that performs important pragmatic and social functions. For instance, humor allows speakers to communicate potentially sensitive or difficult topics and also serves to help build relationships (Martin 2007), both of which promote more successful conversations. Humor can also play an important role in creativity, especially during collaborative creative sessions, because humor can prompt more conversation and greater idea production (Cundall 2007).

A number of previous studies outside the creativity domain have investigated the linguistic features of humor using NLP approaches that classify humorous from non-humorous texts based on linguistic indices. For example, Skalicky and Crossley (2015) found that humorous Amazon.com product reviews contained significantly higher levels of negative emotion words and significantly lower levels of lexical sophistication when compared to non-humorous reviews. However, no consistent linguistic features of humor have yet emerged from research in computational humor detection because each corpus of humor tends to produce its own linguistic profile (Skalicky et al., 2016a).

Elaboration/Persuasion

Another important marker of creativity is the ability to elaborate on ideas (Kaufman et al. 2008). Elaboration is generally defined as the expansion of ideas already produced. However, during collaborative dialogue, participants may disagree on ideas and solutions to a task and then attempt to persuade one another of the merits of an idea by expanding on the positive and/or negative qualities of that idea (Bench-Capon 2003).

Linguistically, elaboration and persuasion are marked by higher levels of coherence and cohesion (Conner and Lauer 1985). In other words, elaboration and persuasion involve repetition of ideas and lexical items. In recent NLP studies, cohesion was found to be a marker of elaboration scores produced during a collaborative problem-solving task (Skalicky et al. 2016b).

Current Study

In order to gain a better understanding of the linguistic features of creativity, humor, and persuasion, we investigate these language phenomena using NLP tools. Our research questions are as follows:

Research Question 1. Can linguistic features be used to distinguish humorous and persuasive language in the context of a problem solving task?

Research Question 2. Can those same linguistic features along with incidence counts for humor and persuasion based on human ratings be used to predict creativity and elaboration?

Analysis 1

Method

Analysis 1 investigated whether linguistic features can automatically classify humor, persuasion, and other language types produced by participants during a divergent thinking problem solving task.

Creativity Corpus

Creativity data was gathered by asking 38 pairs of undergraduate and graduate students who were financially compensated to complete three separate tasks designed to elicit creative responses. Each of the three tasks presented participants with an open ended task and asked them to generate as many creative solutions as they could for that task. Participants were otherwise unaware of how their answers would be evaluated. For example, one task asked participants to develop solutions to prevent ice accumulation on an unmanned military antenna located in an arctic climate without increasing the weight of the antenna. The other two tasks involved redesigning a baby chair to be adjustable and cutting a flexible rubber pipe. Communication between the participants was computer mediated using a chat program and these sessions comprised the protocols that were subsequently analyzed.

Humor and Persuasion Scores

In order to calculate incidence scores for humor and persuasion, the protocols from each pair were first separated by individual participant turns during each conversation. Each turn was comprised of everything one participant typed until the point where the other participant typed something. Thus, a single turn could span more than one sentence or utterance. The turns were then analyzed by two trained human raters who assigned each turn to one of the following categories: persuasion (when participants challenged each other regarding potential solutions to the creative task), humor (comments eliciting laughter by the other participant or interpreted as potentially humorous by the raters), and other (all other conversation). After the two raters assigned their codes, any disagreements were adjudicated by a third expert rater.

For each dyad, a single text file containing all turns associated with a category was created, resulting in separate files containing all of the *humor*, *persuasion*, or *other* language produced by each dyad. Text files containing fewer than 25 words were removed to ensure enough linguistic coverage in the text for the linguistic analysis. The final corpus consisted of 354 texts (105 humor, 95 persuasion, and 154 other).

Linguistic Analysis

The linguistic features of the texts for each category were measured using three text analysis tools that measure lexical sophistication (TAALES; Kyle and Crossley 2015), discourse cohesion (TAACO; Crossley, Kyle, and McNamara 2015), and sentiment (SEANCE; Crossley, Kyle, and McNamara 2016). TAALES includes over 100 indices related to lexical sophistication (i.e., the complexity and diversity of a speaker's language), such as word and n-gram frequency, lexical diversity, and range. TAACO includes over 150 indices related to cohesion (i.e., degree of meaning overlap in a text) at both the local (i.e., word-to-word and sentence-to-sentence) and global (i.e., paragraph-to-paragraph) levels. SEANCE measures sentiment in text (e.g., positive and negative valence) using over 250 indices that are subdivided based on negation and part of speech.

Statistical Analysis

Before fitting any statistical models to the data, we checked the reported measurements for each variable from all three tools for normality. We then ran Multivariate Analyses of Variance (MANOVA) to determine which of the remaining indices were significant predictors of the three text categories (i.e., humor, persuasion, and other). All indices flagged as significant in the MANOVA were then checked for multicollinearity with one another. For any two indices that demonstrated strong multicollinearity ($r > .89$), we retained the index with the strongest effect size in the MANOVA results. We followed up the MANOVA with a stepwise discriminant function analysis (DFA) to predict category membership. We first trained our DFA on the entire data set and then performed cross-validation using leave-one-out-cross-validation (LOOCV).

Analysis 1 Results

MANOVA

A one-way MANOVA was conducted to examine if the indices from the text analysis tools differed significantly among the three text groups. The MANOVA returned a significant result ($F[354, 352] = 1.32$, $p < .001$) and indicated that 32 of the linguistic indices reported significant effects as a function of text category. The significant indices were related to lexical sophistication (TAALES) and discourse cohesion (TAACO), and indicated that humorous texts contained words with higher imagability and meaningfulness, lower spoken and written word frequency, and lower age of acquisition. Persuasive texts were marked by words with higher spoken and written frequency and higher measures of word-to-word and paragraph-to-paragraph cohesion.

DFA

The stepwise DFA chose four variables from among the 32 indices as the best classifier set for the three different text categories: *Type-Token Ratio: Function Words, Average* (FTTR), *BNC Spoken Word Frequency: All Words Log* (BNC_SWF), *Mean Age of Acquisition: Content Words* (AOA), and *Number of Function Words* (NFW; see Table 1 for descriptive statistics). Using these four indices, the DFA was able to correctly classify 50.3% of the texts, which was significantly better than the 33.3% chance ($\chi2$ [4, n=354] = 38.856, $p < .001$). For the LOOCV, the DFA reported an accuracy of 49.7%. The measure of agreement between actual and assigned text type produced a Cohen's Kappa of .232, representing a fair agreement (Viera and Garrett 2005; see Table 2 for results).

Index	**Humor**	**Persuasion**	**Other**
FTTR	0.525 (0.246)	0.508 (0.209)	0.419 (0.217)
NFW	0.563 (0.074)	0.600 (0.072)	0.573 (0.063)
AOA	5.422 (0.482)	5.648 (0.538)	5.574 (0.488)
BNC_SWF	4.976 (0.210)	5.045 (0.185)	5.041 (0.168)

Table 1: Means and standard deviations of the linguistic indices selected by DFA as significant predictors.

Actual Type	**Predicted Type**			
Total Set	Humor	Persuasion	Other	Total
Humor	38	25	42	105
Persuasion	18	47	30	95
Other	27	34	93	154
LOOCV Set	Humor	Persuasion	Other	Total
Humor	38	25	42	105
Persuasion	19	46	30	95
Other	28	34	92	154

Table 2: Confusion matrix results for DFA trained on whole set and cross-validated set.

Analysis 1 Discussion

Linguistic Features

Results from the DFA suggest that the three text categories (humor, persuasion, and other) can be distinguished with 50% accuracy using four linguistic features. Specifically, both the type-token ratio of function words and the total number of function words emerged as significant predictors, as did mean age-of-acquisition for words and spoken word frequency of content words. Type-token ratio is a measure of the variation in words within a text, and the specific measure here focused on the variation in use of function words (i.e., words that contain little semantic meaning and serve to primarily perform grammatical functions in text, such as *and, the,* and *or*). *Other* texts contained a lower type-token ratio of function words compared to both *humor* and *persuasion* texts (see Table 1), suggesting that *other* texts contained less variability in the types of function words used. Pairwise comparisons confirmed that *other* texts differed significantly from both *humor* (p = < .001) and *persuasion* (p = .003), with no significant difference between *humor* and *persuasion* (p = .586).

In addition, the total number of function words also emerged as a significant predictor variable. As shown in Table 1, *persuasion* contained the highest number of function words and *humor* contained the fewest. Pairwise comparisons confirmed that *humor* and *other* contained significantly fewer function words when compared to *persuasion* (*humor* p < .001, *other* p = .004), whereas there was no significant difference between *humor* and *other* (p = .248).

The third selected index was age of acquisition for content words (i.e., words that contain semantic meaning, such as nouns and verbs). For this index, a lower mean indicates that the text is comprised of words judged to be learned earlier by native speakers. As shown in Table 1, *humor* contained the lowest average with no differences between *persuasion* and *other*. Pairwise comparisons indicated that *humor* differed significantly from both *persuasion* (p = .002) and *other* (p = .017), and that *persuasion* and *other* did not differ significantly (p = .261).

Finally, the fourth significant index was based on spoken word frequency of content words found in the British National Corpus, a large corpus of British English collected during the late 20th century. In general, lower word frequency correlates with the use of more sophisticated vocabulary. The values in Table 1 suggest that *humor* contained lower word frequency, with *persuasion* and *other* equally higher. Pairwise comparisons confirmed significant differences for *humor* when compared to both *persuasion* (p = .019) and *other* (p = .043) with no significant differences between *persuasion* and *other* (p = .553).

Accuracy

In both the full and LOOCV DFA models, classification accuracy for *other* texts was highest (~60%), with *persuasion* the second highest (~49%), and *humor* the lowest (~36%). Spontaneous humor has traditionally been difficult to classify using linguistic indices (Skalicky et al. 2016a), and the results here attest to that challenge. In addition, the lower accuracy for *humor* texts may in part be explained by how the texts were classified. Among the signals for humor were written laughter (e.g., haha) and emoticons (e.g., ☺), both of which represent pragma-linguistic markers of humor and are not be measured by the text analysis tools used in this study.

Summary

The results of Analysis 1 suggest that linguistic differences exist among humor, persuasion, and non-humorous and non-persuasive texts (i.e., *other*). Both humorous and persuasive texts contained more variation in their use of function words, but also significantly fewer instances of function words. This suggests that humor and persuasion employ function words in a more specific and purposeful manner when compared to *other* texts. In addition, humorous texts contain words with lower average age-of-acquisition scores as well as significantly lower average spoken word frequency. This suggests that humor employed in the dyad conversations contained relatively more sophisticated language based on word frequency, but less sophistication when based on age-of-acquisition as compared to both *persuasion* and *other* texts.

Analysis 2

Method

The purpose of Analysis 2 was to investigate whether the linguistic features predictive of humorous and persuasive texts in Analysis 1, as well as the incidence counts of humorous and persuasive categorizations based on human ratings, significantly predict human scores of creativity and elaboration from the same creativity corpus.

Creativity and Elaboration Component Scores

We used *Creativity* and *Elaboration* component scores derived from a previous analysis of the same data (Skalicky et al. 2016b) as our predictor variables. In that study, two trained raters evaluated the chat transcripts of each task for each dyad using an analytic rubric designed to measure creativity. The rubric contained seven subscales using six-point interval scales and measured the total number of ideas (ideation), the number of different idea types (flexibility), the originality of the ideas (originality), elaboration of ideas, use of humor, use of metaphor, and use of word play (i.e., linguistic manipulation of word forms and meanings). These ratings were analyzed using principle component analysis in order to develop weighted component scores for each task per dyad. The results led to a *Creativity* component score that included ideation, flexibility, originality, and humor and an *Elaboration* score that included elaboration and metaphor (word play did not load onto either component and was thus excluded from the analysis).

Humor and Persuasion Counts

We included incidence counts for humor and persuasion for each text by calculating the percentage of humorous and persuasive turns categorized by the human raters (and used in Analysis 1) for each text.

Statistical Analysis

We used linear mixed effect models (LME) to examine whether the linguistic indices derived from the results of Analysis 1 were predictive of the *Creativity* and *Elaboration* component scores obtained by Skalicky et al. (2016b). For both of the LME models, we entered the four linguistic indices and the humor and persuasion counts from Analysis 1 as fixed effects and participants and task topic as random effects. We used the *lmer* package *lme4* (Bates et al. 2015) to build our models, the *lmerTest* package (Kuznetsova, Brockhoff, and Christensen 2015) to derive *p*-values from the models, and the *MuMIn* package (Nakagawa, and Schielzeth 2013) to calculate effect sizes. *MuMIn* calculates effect sizes for LME models by reporting two separate R^2 values: marginal and conditional. Marginal R^2 reports the variance explained by just the fixed factors, whereas conditional R^2 explains the variance for both fixed and random factors.

Analysis 2 Results

LME Predicting Creativity

An LME model predicting *Creativity* component scores using the four linguistic indices reported in Analysis 1 and the incidence of humor and persuasion reported significant effects for *Type-Token Ratio: Function Words* (FTTR), *Spoken Word Frequency: Content Words Log (BNC)* (BNC_SWF), and incidence of humor (Humor Percent) (see Table 3). The first two of these indices reported negative coefficients, suggesting that participant protocols including a lower ratio of function words and lower spoken word frequency resulted in higher *Creativity* component scores. Percentage of humor reported a positive coefficient, suggesting higher levels of humor resulted in higher creativity scores. This model reported a marginal R^2 of .422.

Index	Coefficient	S.E.	*t*
(Intercept)	7.775	2.151	3.614*
FTTR	-1.491	0.206	-7.211*
Humor Percent	0.118	0.026	4.510*
BNC_SWF	-6.956	3.020	-2.303*
AOA	-0.436	0.351	-1.241
NFW	-0.345	0.330	-1.046
Persuasion Percent	-0.051	0.057	-0.891

Table 3: LME model predicting *Creativity* component score (* = significant, all $p < .001$).

LME Predicting Elaboration

An LME model predicting *Elaboration* component scores using the four linguistic indices reported in Analysis 1 and the incidence of humor and persuasion found a significant effect for *Type-Token Ratio: Function Words* (FTTR). This index reported a negative coefficient, suggesting that participant protocols including a lower ratio of function word types resulted in higher *Elaboration* component scores (see Table 4). This model reported a marginal R^2 of .267.

Index	Coefficient	S.E.	*t*
(Intercept)	6.766	1.267	5.339*
FTTR	-0.868	0.034	-6.621*
NFW	-0.328	0.190	-1.722
AOA	0.180	0.161	1.115
BNC_SWF	1.641	1.852	0.886
Humor Percent	0.005	0.016	0.325
Persuasion Percent	-0.002	0.034	-0.067

Table 4: LME model predicting *Elaboration* component score (* = significant, all $p < .001$).

Analysis 2 Discussion

Linguistic Features

Results from the LME models indicated only two of the linguistic indices selected as significant predictors by the DFA in Analysis 1 were significant predictors of the *Creativity* component score, and only one was a significant predictor of the *Elaboration* component scores. Both type-token ratio of function words and spoken content word frequency measures predicted *Creativity* scores, whereas only the type-token ratio of function words predicted the *Elaboration* scores. These results suggest that *Creativity* is marked linguistically by less varied usage of function words, as well as lower spoken word frequency. These findings suggest that the language associated with *Creativity* is sophisticated in its use of content words. This finding partially supports previous findings indicating that *Creativity* is predicted by markers of lexical sophistication (Skalicky et al. 2016b). However, the same cannot be said for function words, which reported a negative coefficient with *Creativity* scores suggesting that creativity is predicted by lower variability in the different types of function words employed. Since function words are used to organize sentence structure, a likely interpretation is that creativity depends on less variable structures from which to evolve more complex ideas.

The results for the *Elaboration* component scores were similar in that lower variability of function word types resulted in higher *Elaboration* scores. Previous results indicated that *Elaboration* is marked by higher lexical cohesion and overlap (Skalicky et al. 2016b), and it may be the case here that the lower type-token ratio for function words represents greater overlap of function words between the dyad participants. It may also be the case that elaboration depends on less variable structures similar to creativity.

Humor and Elaboration as Persuasion

We also tested whether the incidence of humorous and persuasive turns were predictive of the *Creativity* or *Elaboration* component scores. Our results demonstrate that greater humor results in higher *Creativity* component scores, which provides additional evidence supporting connections between humor and creativity during collaborative problem solving (Cundal 2007).

Summary

Indices predictive of humor and persuasion were predictive of scores of creativity and elaboration within the same creativity corpus. Greater creativity was marked by lower word frequency and less variability in function word use, whereas greater elaboration was marked by less function word variability. In addition, higher amounts of humor resulted in higher creativity scores, suggesting that humor aids in the creative process. The marginal R^2 for the model predicting creativity scores is also relatively strong, indicating that 42% of the variance in *Creativity* scores can be attributed to these linguistic features and the use of humor alone. The R^2 for the model predicting elaboration was also relatively strong, indicating 27% of the variance in *Elaboration* scores can be attributed to one linguistic variable (function word type-token ratio).

Conclusion

We conducted two analyses in order to better understand the linguistic features of creativity, elaboration, as persuasion, and humor. The results from these two studies provide insight into the linguistic nature of humor, persuasion, and creativity. Lexical sophistication and textual cohesion emerged as important linguistic features to consider, while strong links between humor and creativity were also found.

Our findings have important implications for current understanding of creativity. By identifying specific language features that signal creativity, this research provides additional means for measuring the creative potential of individuals, as well as a better understanding of creativity itself. Additionally, identification of the linguistic features of related constructs, such as humor and persuasion, allow for a better understanding of how humor and persuasion may affect creative output during natural conversation.

Future research is needed in order to continue investigating the relation between these language forms and their linguistic features. Specifically, linguistic comparisons between humor and persuasion produced during creative and non-creative tasks would help to further define the similarities and differences between humor, persuasion, and creativity.

References

Acar, S., and Runco, M. A. 2014. Assessing Associative Distance Among Ideas Elicited by Tests of Divergent Thinking. *Creativity Research Journal 26*(2):229-238

Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*,*67*(1):1-48.

Bench-Capon, T.J. 2003. Persuasion in Practical Argument Using Value-Based Argumentation Frameworks. *Journal of Logic and Computation 13*(3):429-448.

Connor, U., and Lauer, J. 1985. Understanding Persuasive Essay Writing: Linguistic/Rhetorical Approach. *Text-Interdisciplinary Journal for the Study of Discourse 5*(4):309-326.

Crossley, S.A. Kyle, K., and McNamara, D.S. 2015. Tool for the Automatic Analysis of Text Cohesion (TAACO): Automatic Assessment of Local, Global, and Text Cohesion. *Behavior Research Methods* 1-11.

Crossley, S.A., Kyle, K., and 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* 1-19.

Crossley, S.A., Muldner, K., and McNamara, D.S. 2016b. Idea Generation in Student Writing: Computational Assessments and Links to Successful Writing. *Written Communication 33*(3):328-354.

Cundall Jr, M.K. 2007. Humor and the Limits of Incongruity. *Creativity Research Journal 19*(2-3):203-211.

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

Kuznetsova, A, Brockhoff, P.B., and Christensen, R.H.B. 2015. lmerTest: Tests in linear mixed effects models. R package version 2.0-29, http://CRAN.R-project.org/package=lmerTest

Kyle, K., and Crossley, S. A. 2015. Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly 49*(4):757–786.

Martin, R.A. 2007. The psychology of humor: An integrative approach. San Diego, CA: Elsevier.

Nakagawa, S., and Schielzeth, H. 2013. A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-Effects Models. *Methods in Ecology and Evolution 4*(2):133-142.

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

Skalicky, S., and Crossley, S.A. 2015. A Statistical Analysis of Satirical Amazon.com Product Reviews. *The European Journal of Humour Research 2*(3):66-85

Skalicky, S., Berger, C.M., Crossley, S.A., and McNamara, D.S. 2016a. Linguistic Features of Humor in Academic Writing. *Advances in Language and Literary Studies 7*(3):248-259.

Skalicky, S., Crossley, S.A., McNamara, and Muldner, K. 2016b. Predicting Creativity in Task-Based Problem Solving Using Linguistic Features. Poster presented at the 26th Annual Meeting of the Society of Text and Discourse. Kassel, Germany.

Tidikis, V., & Ash, I. K. 2013. Working in Dyads and Alone: Examining Process Variables in Solving Insight Problems. *Creativity Research Journal 25*(2):189-198.

Viera, A.J., and Garrett, J.M. 2005. Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine 37*(5):360-363.