Vocabulary Levels Have No Significant Effect on Impressions of Credibility

Phuc H. Le, Charles L. Peppers, and Hannah G. Bentley

Grinnell College

Acknowledgements

We would like to express our sincerest appreciation for Professor Laura M. Sinnett of Grinnell College for assisting us with the experimental design and the data analysis processes. We have no conflict of interest to disclose.

Abstract

Past research has shown that vocabulary levels affect impression formation of others in written forms of communication across different contexts. This study investigated the relationship between vocabulary levels and credibility, a specific characteristic important for forming relationships in daily life. Undergraduate students ($N = 66$) read a short passage adapted from *The New York Times* that reflected one of three vocabulary levels (i.e., low, control, and high). They then completed a 20-item questionnaire that asked about their ratings of the author's credibility. The results indicated no significant difference across the three groups in how participants perceived credibility, suggesting that vocabulary levels did not affect impressions of credibility among college students. Alternative explanations for this null relationship are further discussed with some methodological revisions for future research.

Keywords: impression formation, vocabulary, credibility

Vocabulary Levels Have No Significant Effect on Impressions of Credibility

Whenever we encounter someone - whether they are a close friend we always accompany at weekend parties, a partner we collaborate with in a class project, or a stranger with whom we share a table in a diner - it is difficult to avoid forming impressions of that person. Impressions allow us to make quick judgments about a person or group, adjust our behaviors to fit with our perceptions of these targets, and facilitate meaningful social interactions. Among various types of impressions, credibility plays a significant role in influencing behaviors towards others. An individual's credibility can be expressed through physical appearance, verbal and nonverbal communication, general behavior, and a cue's salience among other things (Smith et al., 2015). In this paper, we defined credibility as the level of confidence an individual had, in that an author of an article was both qualified to write that article and qualified, intelligent, and logical in general. Credibility is especially important when it comes to the role of trust, as well as in general impression formation and everyday relationships. Whether or not the person doing someone's taxes can be trusted to do them well relies on impressions of the tax person's credibility. Be it recommendations from friends or 30 years of service, one has to know if their taxes will be completed correctly.

Credibility in impression formation can be especially important when it comes to the written form, as with applying to jobs, sending emails, and communicating with colleagues. Literature on the impression formation process of credibility has shown the prevalence of how vocabulary level impacts this kind of impression. Stevens (2005) illustrated the real-world impact of vocabulary levels by depicting that employers look for certain types of elevated vocabulary in resumes and cover letters when hiring employees. His research showed an

emphasis on written communication over other types, which is especially conducive to our research's operations. Vocabulary was also the focus of Sun and colleagues' (2019) study. However, the article studied spoken, rather than written, vocabulary, thereby demonstrating the possible differences between the two media. Such differences in presentation form of vocabulary is meaningful to our study because they give us an idea as to how having the subjects simply read the manipulated passage could change our results.

Forming impressions based on the way people express themselves through text is becoming a more salient issue in our technological age. The way a person presents themself using vocabulary and citations affects the way others view them and specifically their credibility, both of which are extremely important, especially in terms of employment. Darbyshire, Kirk, and Kaye (2016) used Facebook to show some traits such as openness and conscientiousness can be accurately assessed based on online text. These two themes are related to intelligence, which overlaps with our concept of credibility. People who are more intelligent are likely to be viewed as more credible. For example, people are more likely to accept tutoring from a [ ] College student than from a fraternity member at University of [ ], because they would likely attribute more intelligence and ability to the former.

Our prediction is based on Chaiken's (1980) persuasion heuristic theory, which states that a cue can make people like or dislike an attitude object without thinking about it in any depth, or by using automatic processing. Reinhard and Sporer (2010) examined how task involvement regulated the use of systematic or superficial processing influences the credibility of a person giving information. In the first experiment, the researchers used consistency, plausibility, and commentator attractiveness to assess whether participants used source information (systematic

processing) or source cue information (superficial processing) to determine credibility. The results indicated that when participants were not closely involved in the task, only commentator attractiveness affected their impressions of credibility. However, participants also took into consideration content information to make inferences about the source's credibility when they actively engaged themselves in what they were doing. Experiment 2 extended the results of Experiment 1 by testing the effects of the different types of processing on specific statements. Experiment 3 examined the effect of verbal content on detection of lying. These three experiments relate to our study because we hope that readers will utilize systematic processing to determine credibility from the vocabulary in the passages they read.

As for data that conflict with our hypothesis, Oppenheimer (2006) investigated the use of long words, rather than specifically higher levels of vocabulary in general. The researcher found that subjects often discounted intelligence when the author used unnecessarily long words. These findings partially refute our hypothesis, which is that higher levels of vocabulary will result in subjects thinking the author of the passage is more credible. However, Oppenheimer's results were based on word length instead of vocabulary level. This difference in design led us to expect higher ratings of credibility among participants in the high vocabulary condition than those in the control group or low condition.

We hypothesized that using more difficult words could lead the reader to perceive greater persuasiveness and thus, higher credibility of the texts and their authors than that of lower-level vocabulary. Our prediction was that participants in the higher vocabulary level condition would have more difficulty in interpreting to the fullest extent the meaning of the passage and thus would engage in superficial processing, which would lead them to judge the article author to be

more credible than those in the normal vocabulary level condition. In contrast, participants in the lower vocabulary level condition, in comparison with their counterparts in the medium (control) vocabulary level condition, would easily process the meaning of the passage, think more systematically, and rate the article author to be less credible.

Method

Participants

The study involved 68 students aged 17 to 22 at [ ] College, a liberal arts college in the Midwest. We had to eliminate two participants, the first of whom made some technical errors and the second failed to read the instructions and follow the procedure. In our final sample ($N = 66$) , there were 20 men and 40 women, as well as 6 non-binary participants. The sample was comprised of 43.94% White, 34.85% Asian, 6.06% Latinx, and 3.03% Black or African American, and the remaining 12.12% were individuals whose had multiple or mixed racial or ethnic identities. We recruited participants by posting a signup sheet on the bulletin board of the department, asking personal contacts outside of the college's dining hall, publishing Facebook posts in the student groups, spreading word of mouth, and sending out emails or texts. Participants in introductory-level psychology courses received 0.5 PSELL credit for their participation. For those not enrolled in Introduction to Psychology at the time of the experiments, the Psychology department provided a drawing for a $20 college bookstore gift card. We had to eliminate two participants, the first of whom made some technical errors and the second failed to read the instructions and follow the procedure.

In order to test whether vocabulary would have an effect on perceived credibility, we manipulated the level of vocabulary in three separate passages, adapted from the same *New York*

Times article (Alhindawi & Katz, 2019). We randomly assigned participants to one of the three conditions – low, control or medium, and high – according to the session in which they participated, and we asked them to rate the credibility of each passage's author.

Materials

An article from *the New York Times* by Alhindawi and Katz (2019), "The Disappearing Schools of Puerto Rico," was selected for testing for two reasons. Firstly, *The New York Times* is one of the most common and available newspapers for [ ] professors and students. Moreover, the chosen article topic was relevant to the academic setting: it discussed an educational issue concerning the closure of hundreds of schools in Puerto Rico. The chosen excerpt had three paragraphs, describing the situation of Puerto Rican schools, US "aid," and the effect of US interference on education. In each paragraph, we changed from two to three words (e.g., nouns, verbs, and adjectives) to their more advanced or basic synonyms in order to match higher and lower vocabulary levels, respectively, making for seven changed words in all for each condition. In each condition, the same words were changed. Google's thesaurus function was our reference for these synonyms, and we evaluated the complexity of vocabulary level on word length and commonality. For example, the words "fleeing" and "embarked" in the medium or control condition became "loss" and "started" in the lower vocabulary level, while in the higher vocabulary level, they were "diaspora" and "inaugurated" respectively.

The ratings were on a 7-point Likert scale and consisted of 20 questions to assess the participants' impressions about the author of the paper. Because impressions of credibility were the focus of the study, there were 12 questions (60%) targeting specifically the author's perceived credibility, including statements regarding the author's qualification, occupation,

trustworthiness, knowledge of the situation, worthiness for citation, and ability to do research. For example, we asked participants the following question: “How likely are you to cite this author if you were writing a paper on the state of affairs in Puerto Rico?” Other questions that did not specifically have to do with the hypothesis pertained to the author’s perceived levels of friendliness and carefulness. For example, we asked the following question: “How invested in Puerto Rico’s recovery do you think the author of this passage is?” We included questions like these in order to make sure the subjects would not figure out exactly what the survey was measuring, which could have potentially compromised the data’s validity. In addition, we separately analyzed data from some of these statements which did not directly address the research question.

Procedure

We ran our experiments on groups of one to six participants at a wide range of times throughout a period of three weeks in a college computer lab. Once participants completed the consent form, they received the printed passage to read and then the instructions to complete the ratings on an online survey after finishing reading. We decided to print out the text instead of presenting it online because a paper copy would be most accessible for participants’ reading abilities, allow for easy note-taking while they were reading, and cause less fatigue for the eyes upon reading multiple times. We considered the online survey superior to its offline counterpart because the former streamlined the data entry process and increased the anonymity of data by avoiding handwritten texts. After participants finished the survey, they received a debriefing form about the purpose of the study before exiting the room.

Results

Scoring

We reverse-scored answers to questions that directly measured the participants' perceived credibility of the author (i.e., items 2, 3, 4, 9, 12, and 16). We had negatively worded these specific items to eliminate acquiescence response bias among participants who rushed to complete the survey. After reverse-scoring these items, we aggregated the means across the survey questions about credibility in order to obtain the means across conditions. Included in this aggregate were items 1, 8, 10, 13, 17, 20, and their reverse-scored counterparts.

Analysis

We conducted a one-way ANOVA statistical test because we did not analyze gender or race due to small sample sizes in each group. Additionally, we conducted an a priori contrast tests in order to give ourselves a liberal range of error. When we looked at variables individually, outside of the aggregate mean score, all *F*s were less than .98, and all *p*s were greater than .09, but neither of those values came from questions pertaining directly to our hypothesis. For the questions pertaining to our hypothesis,we found that all *F*s were less than .97, and all *p*s were greater than .39, showing that our manipulations of credibility yielded no significance across the three conditions (see Figure 1). An a priori contrast test (low = −1, control = 0, high = 1) also yielded no significance, $t(2, 63) = -.30$, $p = .77$. Changing the scoring for the test (low = 0, control = 1, high = −1) also yielded no significance, $t(2,63) = .12$, $p = .90$. For all three conditions, the means were very similar ($M_{low} = 4.56$, $SE = .26$,$M_{control} = 4.51$, $SE = .28$, $M_{high} = 4.44$, $SE = .29$), as expected, given the lack of significance.

However, there was a marginally significant difference across conditions with a question unrelated to our hypothesis: How involved in the Puerto Rican crisis does the author seem to be, $F(2, 63) = 2.45$, $p = .09$. An a priori contrast test (low = −1, control = 0, high = 1) yielded significance, $t(2, 63) = -2.07$, $p = .04$, $M_{low} = 5.00$, $SE = .33$, $M_{control} = 4.57$, $SE = .37$, $M_{high} = 4.52$, $SE = .43$. However, we believe this is due to random error rather than any real relationship between the question and conditions, because the item was not related to our hypothesis, and it had only marginal significance. We therefore found no significance for any statistical tests, and thus we failed to reject the null hypothesis of there being no difference across conditions.

Discussion

We hypothesized that advanced-level vocabulary would make participants judge the article author to be more credible than basic-level vocabulary because the former requires more mental resources to understand, impedes the smooth flow of information processed, thus allowing for superficial processing. Along with this, we expected that an intermediate level of vocabulary would result in an intermediate level of credibility (potentially an average score of 3 or 4 on a 7-point scale) associated with the author because of the use of slightly complex words to which most college-level readers are used. We were unable to accept this hypothesis due to the data our study produced. We found no significant differences across levels of vocabulary for any of the questions on our survey. However, we did find marginal significance with one survey question, though this was likely due to random chance. We should also note that we did find a very small trend in lower mean scores being associated with the lower vocabulary level condition. However, this trend was obviously not significant.

Other studies, such as the one by Darbyshire, Kirk, and Kaye (2016), show that people are fairly accurate when assessing traits of online targets, but our study did not reflect this finding. This is likely because the vocabulary level manipulation was not strong enough. Further, the author of the passage was always the same, and participants may have just assumed that the author was of intermediate credibility. Oppenheimer (2006) found that the use of long words caused subjects to often discount the author's intelligence. It is unlikely that our high-level vocabulary words (e.g. exodus, inaugurated) were on par with the words the researcher used (i.e. the title of the paper was "Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly," illustrating the researcher's point). Thus, while students may have discounted the words as unnecessarily complex or technical in Oppenheimer's study, our study may not have produced the same perceived difficulty in processing that engages superficial processing.

A source of strong construct coverage is the different facets we included to determine and define credibility. We operationalized credibility in six different ways: perceived author's trustworthiness, knowledgeability, citability, ability to do research, occupation, and qualification. By covering a wide spectrum of credibility, we are confident that we controlled for participants' different interpretations of this quality and that they found no ambiguity concerning "credibility." This being said, a potential explanation for the lack of significant results could be explained by the vocabulary levels not being different enough or unique enough to make a difference in credibility ratings. We should have also included more levels of vocabulary to diversify the words participants encountered. For example, we could have had very high and very low levels, in addition to our high, low, and intermediate vocabulary levels. Additionally, we could have

manipulated more words in the passage itself in order to make the difference in manipulations stronger. Since we only doctored seven words from the original paragraph for this study, these words may have not accurately expressed vocabulary levels. We were unable to validate the effectiveness of our manipulation because we did not conduct a manipulation check to test for participants' perceptions of different vocabulary levels in the passage. However, conducting a manipulation check would have compromised our cover story. Participants would attend more closely to words that, because of our manipulation, appear longer and less common, and that do not fit with the context. As a result, they could recognize the untold purpose of our study and even the conditions they are in, and their behaviors and performance could reflect the Hawthorne effect. Moreover, observations of participants' individual differences in terms of how they read and marked up the passage during the experiments inform us that manipulating only vocabulary levels could not fully capture the effects of vocabulary levels. Thus, we should include other factors at play. To more accurately measure these effects, researchers in future studies could include a scale variable and a nominal one: the amount of time it takes participants in different conditions to complete the ratings and whether they go back and forth between the passage and the ratings during the experiments, respectively. Both would need to be added due to our observations of individual differences with subjects taking different amounts of time and using different reading and annotation methods. Both variables would act as covariates, which, when taken into account, could help to control for what the researcher is actually trying to measure. Additionally, student participants often rush to complete surveys and thus spend less time than is recommended or needed to carefully consider the content of the passage and the questions. Though it is difficult to know how long it would take each individual to comprehend the words

and respond to the survey, it is likely that students superficially processed the passage's content not because of the vocabulary, but because of lack of attention to the passage itself. Therefore, construct validity is not the best due to the weakness of differences between vocabulary levels, and internal validity is not high because of limited engagement from participants.

The student body of [ ] College is not representative of the U.S, and because we used [ ] students for subjects, our sample was drawn from a very limited population. Therefore, our subjects may limit the generalizability (and external validity) of our study. We would expect a more advanced vocabulary from [ ] students than from students at a large, public university because of the selectivity of the college. To be accepted into [ ] College, students must be academically advanced or at the very least competent, meaning they often have more impressive academic capabilities or skill sets than those from less selective colleges and universities. Therefore, it is very likely that their vocabulary levels may be higher than those of an average college student or person in the United States. Further, [ ] students tend to take classes which inform them of the type of phenomenon which we were testing, making them possibly less susceptible to the manipulations.

Concerning the demographics of the sample, we recognize that there are a disproportionate number of non-binary people. In addition, women account for more than half of all participants. This gender imbalance does not allow us to test for gender effects and generalize our results to the general population. We also acknowledge that participants who identify themselves as White or Asian account for 78.79% of the sample, while people of Black or African American and Hispanic or Latinx origins take up less than 10% of the sample. This could have resulted from our sampling of the student population at [ ] College. Besides the

limited number of participants who signed up through the bulletin board, we had to recruit more participants from our personal connections. This convenience sampling makes the majority of these individuals limited to domestic Caucasian students and international students of South Asian, East Asian, and South East Asian descents, instead of extending to other ethnic groups.

Future studies should conduct testing in advance of the conditions before the actual experiment. Researchers could run a pretest with at least 20 participants in each condition of vocabulary levels. Conducting this manipulation check would ensure that the vocabulary levels are accurately rated based on intended difficulty or weakness. Participants would read a longer passage adapted from *The New York Times* in a limited amount of time and rate their comprehension of the text. A longer passage would allow for more words to be changed to reflect the intended vocabulary level, thus strengthening the manipulation. Further, researchers could take a passage from a source other than *The New York Times*, such as *CNN*. This could provide for more construct validity, as *The New York Times* and other sources like it typically already use higher vocabulary levels. By starting with a source which employs more average vocabulary levels, the study could gain more construct validity overall. Additionally, researchers could also resort to more qualified resources other than Google's thesaurus, such as members who work at the Writing Lab at their college. These people have undergone rigorous training and thus are sufficiently qualified to help with different aspects of writing, which include vocabulary. Because of their qualifications, they can provide guidance on how to manipulate the vocabulary levels more accurately. With these changes, we hope that we could find support for our original hypothesis that credibility rating is positively correlated with vocabulary level.

If these changes to the experimental design find significance, there are many ways to extend the literature surrounding vocabulary and credibility. First, it would be interesting to conduct a comparison study between elite college-educated and non-college-educated samples. It would make sense that if our prediction that students elite institutions have higher-than-average vocabulary is correct, then they would be less persuaded by academic jargon and would therefore rate authors' passages as less credible than those without a college education. People without a college education may rate credibility higher because they engage in superficial processing due to more limited vocabulary. The procedure would be basically identical to our study or any revision of it, but with the addition of a subject variable: education level.

Additionally, using a less accessible topic than school closure in Puerto Rico may be crucial in finding significant differences between vocabulary levels. In her dissertation on the effect of vocabulary on teaching introductory microbiology, Richter (2011) predicted that the use of Anglicized English to teach biology would result in better performance than if classical (and more difficult) terms were used. This was found to be the case, showing evidence that more easily understandable and accessible vocabulary allows students to learn better and be more successful in class. Since we assume that longer, more technical terminology leads to superficial processing, it would make sense to choose something like neuroscience or organic chemistry to increase the probability of participants engaging in superficial processing. Therefore, we would also change the topic of the passage we use in the study to something more technical or not easily understandable. Both of these concepts reflect the idea that the vocabulary and topic need to be unfamiliar to participants in order to induce superficial processing, not just one or the other.

Conclusion

We were unable to reject the null hypothesis because our results did not indicate that using more difficult words leads the reader to perceive greater persuasiveness and, thus, higher credibility of the texts and of their authors than that of lower-level vocabulary. However, previous research on vocabulary levels and impression formation has been able to reject similar null hypotheses regarding vocabulary levels and impression formation. We have found possible reasons why this study may differ from others, including the population used and the levels of manipulation. Thus, future research should involve more extreme manipulations of vocabulary levels, different populations, and greater internal and external validity overall.

References

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752–766. http://dx.doi.org/10.1037/0022-3514.39.5.752

Darbyshire, D., Kirk, C., Wall, H., & Kaye, L. (2016). Don't judge a (face)book by its cover: Exploring judgement accuracy of others' personality on Facebook. *Computers in Human Behavior, 58,* 380–387. http://dx.doi.org/10.1016/j.chb.2016.01.021

Goodwin, C. J., & Goodwin, K. A. (2013). *Research in psychology: Methods and design* (7th ed.). Hoboken, NJ: Wiley.

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology, 20*, 139–156. http://dx.doi.org/10.1002/acp.1178

Reinhard, M., & Sporer, S. L. (2010). Content versus source cue information as a basis for credibility judgments: The impact of task involvement. *Social Psychology, 41*, 93–104. http://dx.doi.org/10.1027/1864-9335/a000014

Richter, E. (2011). The effect of vocabulary on introductory microbiology instruction (Order No. AAI3449940). Available from PsycINFO. (921293926; 2011-99230-039).

Smith, E. R., Mackie, D. M., & Claypool, H. M. (2015). *Social psychology* (4th ed.). New York, NY: Psychology Press.

Stevens, B. (2005). What communication skills do employers want? Silicon Valley recruiters respond. *Journal of Employment Counseling, 42*(1), 2–9. https://doi.org/10.1002/j.2161-1920.2005.tb00893.x

Sun, J., Schwartz, H. A., Son, Y., Kern, M. L., & Vazire, S. (2019). The language of well-being: Tracking fluctuations in emotion experience through everyday speech. *Journal of Personality and Social Psychology, 118*(2), 364–387. https://doi.org/10.1037/pspp0000244

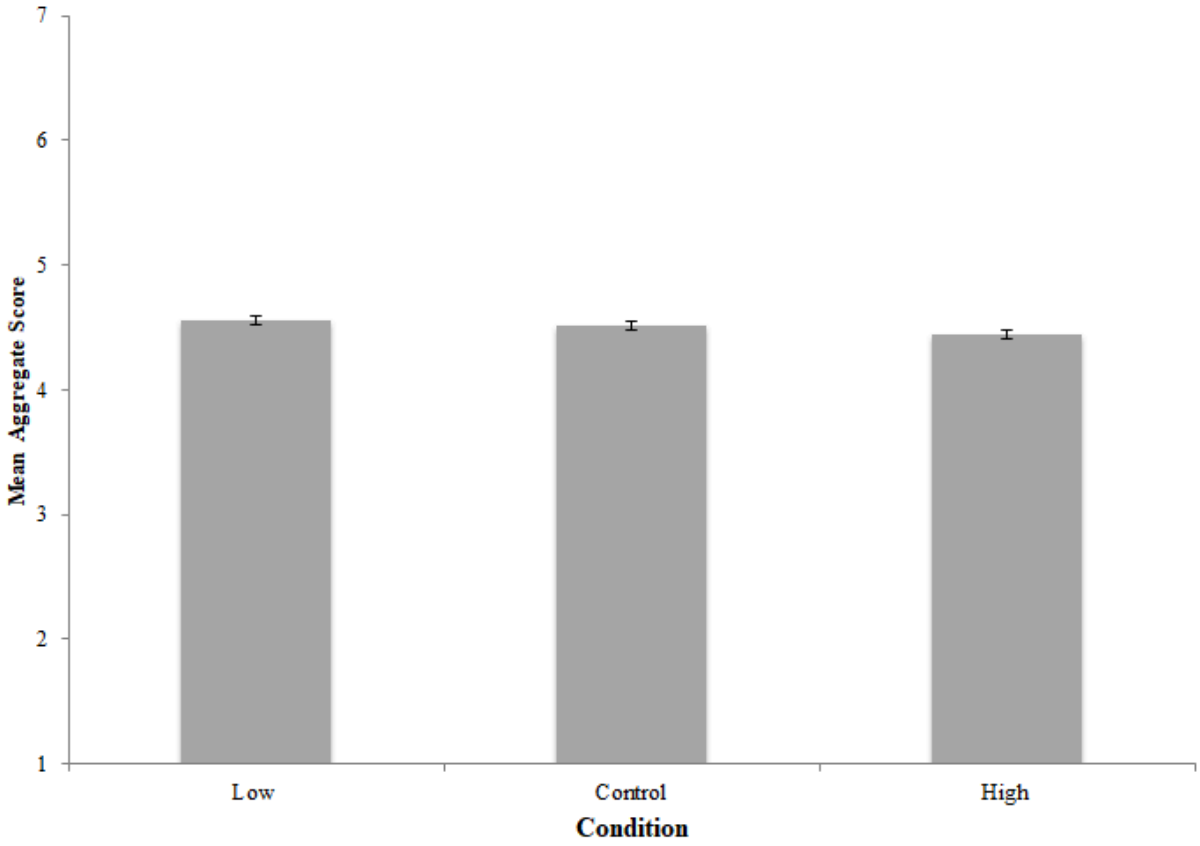

Figure 1. Mean aggregate score of the survey questions by condition. Error bars represent standard error.