

Does Topic Matter? Topic Influences on Linguistic and Rubric-Based Evaluation of Writing

Nia Dowell, Sidney D’Mello, Caitlin Mills, Art Graesser

¹Department of Psychology, Institute for Intelligent Systems, The University of Memphis,
Memphis TN 38152 USA
{ndowell, sdmello, cmills2, graesser}@memphis.edu

Abstract. Although writing is an integral part of education, there is limited knowledge on how assigned topics influence writing quality both in terms of micro-level linguistic features and macro-level subjective evaluations by human judges. We addressed this question by conducting a study in which 44 students wrote short essays on three different topics: traditional *academic-based* topics such as the ones used in standardized tests, *personal emotional experiences*, and *socially charged* topics. The essays were automatically scored on five linguistic dimensions (*narrativity*, *situation model cohesion*, *referential cohesion*, *syntactic complexity*, and *word abstractness*). They were also manually scored by human judges based on a rubric focusing on macro-level dimensions (i.e., introduction, thesis, and conclusion). The results indicated that topic-related differences were observed on both the rubric-based and linguistic assessments, although there were weak relationships between these two measures.

Keywords: Writing quality; Linguistics; Coherence; Coh-Metrix; Cohesion

1 Introduction

Considering the high stakes placed on writing competency in the 21st century, it is not surprising that computational systems utilizing natural language processing techniques have been developed to automatically score written essays and provide interventions to promote writing proficiency (e.g., Intelligent Essay Grader, E-Rater, Summary Street, and Writing Pal). However, little is known about what factors influence the quality of writing; an area that could potentially benefit the advancement of such systems.

Some research has demonstrated that writing quality may be influenced by the topic the individual is writing about [e.g., 1]. A satisfactory understanding of topic influences on writing quality is necessary to ensure that automated writing interventions are optimally beneficial to students. The present research addressed this issue by examining the degree to which both linguistic features and rubric-based assessment scores vary as a function of essay topic. We collected a corpus of essays on three topics and scored the essays using a holistic rubric and Coh-Metrix, an automated text analysis tool that evaluates texts on a number of dimensions [2].

2 Methods

The participants were 44 undergraduates who participated for course credit. The study had a within-subjects design in which the participants were asked to write essays on three topics, namely *socially charged issues* (e.g., abortion, death penalty), *personal emotional experiences* (e.g., write about a happy experience), and *traditional academic prompts* (e.g., debates about extending high school) similar to ones a student might encounter on standardized tests. Within each topic, participants were presented with a number of subtopics and were asked to write for 10 minutes on a subtopic of their choice. A computer interface was used to facilitate typing of the essays. Texts from the 132 essays were saved for offline analyses.

Computational Evaluation. The following is a description of the five primary Coh-Metrix 2.0 [2] dimensions that were used to automatically score the essays. *Narrativity breakdowns* refer to deviations from a sequence of episodes with actions and events that convey a story. *Situation model cohesion* and *referential cohesion breakdowns* occur when there are problems associated with text that are not cohesively connected at a deeper conceptual level or have little overlap in words and ideas, respectively. *Syntactic complexity* refers to structurally dense and embedded sentences that are difficult to process. Finally, *word abstractness* pertains to the extent to which the text contains abstract words (e.g., democracy) compared to words that are more concrete (e.g., table). It should be noted that the Coh-Metrix measures refer to textual problems, so higher numbers indicate either breakdowns in particular dimensions, more complexity, or greater abstractness. It is hypothesized that an essay that is clear should score lower on all these dimensions.

Human Evaluation. Two trained raters (interrater reliability $r = 0.9$) evaluated the essays using a holistic rubric [3], which is similar to the standardized rubric used in assessing essays on the SAT. The overall score was on a 6-point scale with a score of 1 indicating little or no mastery and a 6 indicating clear and consistent mastery. Note that the scores were standardized among each judge to remove any potential bias.

3 Results and Discussion

A repeated-measures MANOVA was performed to investigate the effect of topic on the five Coh-Metrix dimensions. The analysis revealed there was a significant main effect for essay topic, $F(2, 86) = 11.08, p < .001$. Posthoc tests with Bonferroni correction were conducted to identify significant ($p < .05$ for all analyses unless specified otherwise) differences across topics.

The results indicated that students' *academic* essays ($M = -.68, SD = .70$) had the highest frequency of *narrativity breakdowns*, when compared to *socially charged* ($M = -.98, SD = .85$) and *personal emotional experience* essays ($M = -1.8, SD = .88$). However, *academic* essays contained less referential cohesion breakdowns ($M = -.69, SD = .89$) when compared to the *socially charged* essays ($M = -.27, SD = .75$).

Students' *personal emotional experience* essays were characterized by story-like features (less *narrativity breakdowns*). However, these essays were also accompanied by more complex syntax ($M = 1.0, SD = .70$) than the *socially charged* ($M = .62, SD =$

.74) and *academic* essays ($M = .67, SD = .83$). Students also used significantly more concrete words when writing about a *personal emotional experience* ($M = .17, SD = 1.0$) compared to a *socially charged* topic ($M = 1.1, SD = .85$).

Essays on *socially charged* topics ($M = -.98, SD = .85$) had less narrative-like features than essays on *personal emotional experiences* ($M = -1.8, SD = .88$). *Socially charged* essays were also characterized by more abstract words ($M = 1.1, SD = .85$) than essays on both *personal emotional experiences* ($M = .17, SD = 1.0$) and *academic* topics ($M = .47, SD = .69$).

An ANOVA on the rater-provided essay scores indicated that overall scores varied as a function of topic $F(2, 84) = 8.23, MSE = .398, p < .001$. Posthoc tests indicated that the *socially charged* essays ($M = -.32, SD = .88$) were rated lower than the *academic* essays ($M = .16, SD = .95$) and the *personal emotional experience* essays ($M = .15, SD = .94$), which were on par with each other.

We examined the relationship between the two different measures of essay quality by computing a 5×3 (Coh-Metrix measure \times topic) matrix with each cell representing the Pearson's correlation between a linguistic feature and a rubric-based score for a particular topic. The mean absolute correlation was .14, which signifies a small relationship between linguistic and rubric-based evaluations.

4 Conclusions

The results presented here indicate that essay topic can have an impact on writing quality, in terms of both the micro-level linguistic features as well as the more macro-level rubric-based assessments. In line with this, computational systems aiming to advance students writing proficiency can undoubtedly benefit from taking into account such topic-related writing influences.

Acknowledgments. This research was supported by the National Science Foundation (ITR 0325428, HCC 0834847, BCS 0904909, DRK-12-0918409), the Institute of Education Sciences (R305G020018, R305A080589), The Gates Foundation, and START (DHSZ934002/UTAA08-063). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

References

- [1] Beers, S. F., Nagy, W.: Syntactic Complexity as a Predictor of Adolescent Writing Quality: Which Measures? Which Genre?. *Reading and Writing: An Interdisciplinary Journal*. 22, 185--200 (2009)
- [2] Graesser, A.C., McNamara, D.S., Louwerse, M.M., Cai, Z.: Coh-Metrix: Analysis of Text on Cohesion and Language. *Behavior Research Methods, Instruments, and Computers*. 36, 193--202 (2004)
- [3] McNamara, D.S., Crossley, S.A., McCarthy, P.M.: Linguistic Features of Writing Quality. *Written Communication*. 27, 57--86 (2010)