Paper Title The Influence of Essay Prompt Directedness on the Content Quality of Summary Writing

Author(s) Roy B. Clariana, Pennsylvania State University; Ryan L. Solnosky, Pennsylvania State University

Session Title Division C Section 1a Literacy Paper Session

Session Type Paper

Presentation Date 4/14/2023

Presentation Location Chicago, IL

Descriptors Faculty Issues, Writing

Methodology Quantitative

Unit Division C - Learning and Instruction

DOI https://doi.org/10.3102/2004181

The Influence of Essay Prompt Directedness on the Content Quality of Summary Writing

by Roy B. Clariana (rbc4@psu.edu) and Ryan Solnosky, Pennsylvania State University

Abstract:

As an important step in the development of a browser-based writing-to-learn software that provides immediate structural feedback, we seek ways to improve the quality of students essays and to optimize the software analysis algorithm. This *quasi-experimental* investigation compares the quality of students' summary writing under three writing prompt conditions, identical prompts add either 0, 14, or 26 key terms. Results show that key terms matter substantially – the summary essays of those given the prompt *without key terms* had longer essays and the resulting networks of those essays were more like the expert referent and also like their peers' essays. Although tentative, these results indicate that writing prompts should NOT include key terms.

1. Objectives or purposes

This investigation considers how the "directedness" of a writing prompt (i.e., providing key terms in the prompt), influences essay quality. This investigation borrows prompt directedness from the concept map literature. Ruiz-Primo (2004) note "*High-directed* concept map tasks provide students with the concepts, connecting lines, linking phrases, and structure. In contrast, in *a low-directed* concept map task, students are free to decide which and how many concepts they include in their maps." (p. 2) Cuing the key terms when writing *should* improve overall essay quality and could also reduce software pattern matching errors.

2. Perspective(s) or theoretical framework

Writing is an important life-skill and a potent way to learn (see meta-analysis by Bangert-Drowns, Hurley, & Wilkinson, 2004). Blevins-Knabe (1987) makes the point that learning to write and writing to learn go hand-in-hand. A meta-analysis by Graham and Perrin (2007) reported the effect sizes of 11 writing interventions, in order: strategy instruction (0.82), *summarization (0.82)*, peer assistance (0.75), setting product goals (0.70), word processing (0.55), sentence combining (0.50), inquiry (0.32), prewriting activities (0.32), process writing approach (0.32), study of models (0.25), grammar instruction (– 0.32). Summary writing that is used in this present investigation is near the top of their list.

The Writing across the Curriculum Clearinghouse (WACC, 2022) defines writing-to-learn activities as short, impromptu or otherwise informal and low-stakes writing tasks that help students think through and summarize the key concepts or ideas presented in a lesson. Often, these writing tasks are limited to about five minutes of class time, or are assigned as brief, out-of-class assignments. But scoring students' summaries is time consuming, especially in large enrollment courses.

Lomask, Baron, Greig, and Harrison (1992) devised a method for scoring student essays based on essay content conceptual structure. As part of the Connecticut statewide assessment of science content knowledge, trained teacher raters manually converted student essays into concept maps and then hand scored the concept maps as a measure of science content knowledge and comprehension. The approach was successful, but the conversion of essays to concept maps is time consuming and subject to human biases (as is all human rater scoring).

Clariana and Koul (2004) developed a software algorithm for converting text to network maps that can then be automatically scored based on similarity to a referent benchmark network (see also Clariana & Wallace, 2007; Clariana, 2010). This tool was developed into an online browser-based tool called Graphical Interface if Knowledge Structure (GIKS). Our interest in this investigation is to further develop GIMS to support learning course content by summarizing it; and specifically in this investigation, how can the writing prompt be improved to optimize the learning benefits of the software. Although the research base is ongoing on how to improve essay writing prompts (Visser, Maaswinkel, Coenders, & McKenny, 2018), unfortunately the writing-to-learn literature is nearly devoid of investigations regarding including key words in the prompt (we continue to search). Thus the need for this investigation.

3. Methods, techniques, or modes of inquiry

This investigation is a quasi-experimental quantitative investigation.

4. Data sources, evidence, objects or materials

Participants – This is a sample of convenience. A recent dissertation by Wang (2021) used this software tool in an undergraduate architectural engineering course and varied the number of key terms included in the prompt, either 14 central key terms (called focused network, sample n = 44), or 26 peripheral key terms (called full network, n = 40). There was an advantage on several of the measures for the focused treatment. Unfortunately a "no terms" control was not included, so a year later in the same course, the same materials, procedure and prompt were used again but with no key terms (n = 73).

Materials – The software presented a writing prompt that stated "*Reflect back on the readings and lecture this week regarding wood-frame building. Write a broad description of this* *in your own words (about 300 words long)*". Then the terms list was placed below it, either 0, 14, or 26 terms listed in alphabetical order, for example: fastener, floor, framing. joists, nails, opening, plate, platform, residential, roof, roof truss, sheathing, stud, and wall. The student essays were saved to the cloud server for later analysis. Play with a GIKS prompt here.

Procedure – During the regularly scheduled course lab time, students logged in to their assigned computer and launched the browser-based writing software and logged in with an assigned 5 character ID generated by the system (e.g., dkjw8). Following the approved IRB, the key that matched student with ID was kept secret by a course assistant, the course instructor was never aware of who participated. The students composed and submitted their lesson summary essay in about 5 to 10 minutes and individually reflected on the immediate feedback provided by the software. Then they continued to do other lab activities.

Measures – Data consists of 157 essays, 84 essays from a previous dissertation and 73 more essays collected later, note that these are the initial essays before immediate feedback was given. This data has not been reported before, all essays were reanalyzed in this investigation using a different approach, specifically the essays were converted to networks based on the 14 central terms and the networks were then compared using the approach describe by Clariana (2010). Network scores were calculated as the percent of overlap between each essay network with a benchmark network (see Figure 1), links in common divided by the average number of links in the two networks.

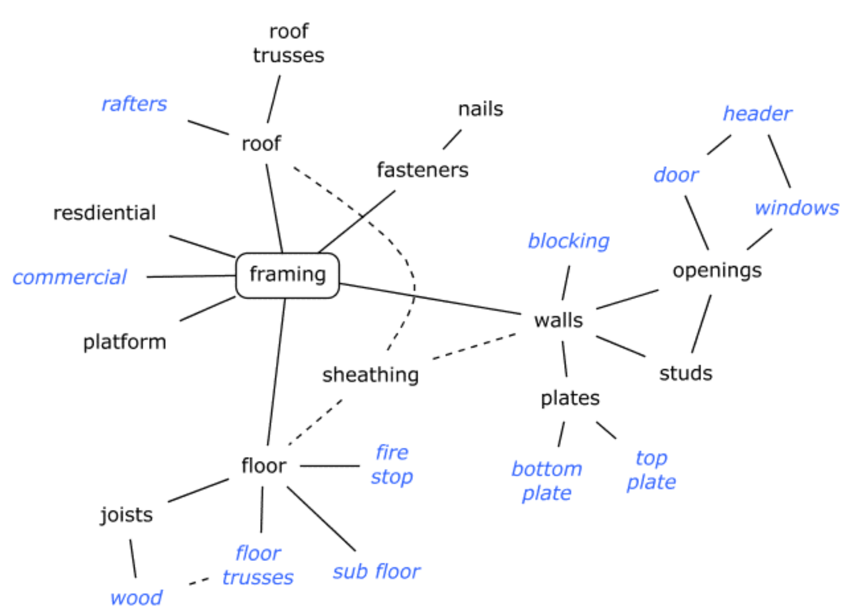

Figure 1. The expert benchmark network used for scoring the students' essay networks. Central terms shown in black and peripheral terms shown in blue.

5. Results and/or substantiated conclusions or warrants for arguments/point of view

The analysis consists of two parts, essay *descriptive measures* (total words, central terms, and peripheral terms) and *network measures* derived from the essays (% overlap with the expert network and % overlap with peers' averaged network, see Table 1).

Table 1. Descriptive and network metrics of the students' essays.

	Descriptive metrics of the essays						Network metrics of the essays			
Treatment	word count		central terms		peripheral terms		% overlap to expert		% overlap to peers	
	M	SD	M	SD	M	SD	M	SD	M	SD
focus (n = 44)	229.3	82.5	8.7	2.9	4.3	2.9	0.36	0.14	0.35	0.13
full (n = 40)	243.8	65.0	8.6	2.4	4.6	2.5	0.40	0.12	0.39	0.10
open (n = 73)	346.3	88.7	8.9	2.2	4.4	2.3	0.44	0.12	0.42	0.11

* *Central terms 14 max, peripheral terms 12 max*

5.1 Descriptive data

Essay length – The total word count in each essay was analyzed by 1-between ANOVA with the factor Prompt Directedness (Focused Full, Open). The data is normal and Levene's test of equality of error variance was *not significant* (Levene statistic (2, 154) = .892, p = .41). Prompt Directedness was *significant*, F(2, 156) = 34.503, MSe = 6584.949, $p < .001$, partial eta squared (η^2) = .309. Scheffe follow-up comparisons show that for total word count, Open > Focus ($p < .001$, ES = 1.21), and Open > Full ($p < .001$, ES = 1.06), adding key terms to the writing prompt decreased the total number of words in the essays.

Central key terms in the essays (14 max) – The number of central terms in each essay was analyzed by 1-between ANOVA with the factor Prompt Directedness (Focused Full, Open). The data is normal and Levene's test was *not significant* (but note: Levene statistic (2, 154) = 2.871, p = .06). Prompt Directedness was not significant, F(2, 156) = .256, MSe = 6.0431, p = .256, η^2 = .003. Including key terms in the writing prompt made little difference on the number of central terms in the essays.

Peripheral key terms in the essays (12 max) – The number of peripheral key terms in each essay was analyzed by a 1-between ANOVA with the factor Prompt Directedness (Focused Full, Open). The data is normal and Levene's test was *not significant* (Levene statistic (2, 154) = 1.186, p = .31). Prompt Directedness was not significant, F(2, 156) = .242, MSe = 6.355, p =

.79, $\eta^2 = .003$. Including key terms in the writing prompt made little difference on the number of peripheral terms in the essays.

Central vs. Peripheral term counts (n = 157) – The number of central and peripheral key terms in each essay were analyzed by paired samples test (2-sided dependent t-test). The simple correlation is r = .63 ($p < .001$). The two measures are significantly different, central (62.7%, M = 8.8 terms) > peripheral (36.7%, M= 4.4 terms), the essays contained substantially more central key terms than peripheral key terms (ES = 1.2).

5.2 Essay Network data

Essay network similarity with the expert network as a measure of essay quality – The similarity of each essay to the expert benchmark referent (as percent overlap) was analyzed by a 1-between ANOVA with the factor Prompt Directedness (Focused Full, Open). The data is normal and Levene's test was *not significant* (Levene statistic (2, 154) = 1.137, $p = .32$). Prompt Directedness was *significant*, F(2, 156) = 5.451, MSe = .017, $p = .005$, $\eta^2 = .066$. Scheffe follow-up comparisons show that for similarity to the expert network, only Open > Focus ($p = .001$, ES = .61), the Full mean fell midway between Focus and Open and was not significantly different than either. Counterintuitively, including only the central key terms in the writing prompt decreased the essay network similarity to the expert.

Essay network similarity with peers' averaged network as a measure of peer-peer knowledge convergence – The similarity of each essay to their peer's averaged network (as percent overlap) was analyzed by a 1-between ANOVA with the factor Prompt Directedness (Focused Full, Open). The data is normal and Levene's test was *not significant* (Levene statistic (2, 154) = 1.270, $p = .28$). Prompt Directedness was *significant*, F(2, 156) = 5.966, MSe = .013, $p = .003$, $\eta^2 = .072$. Scheffe follow-up comparisons show that for similarity to the expert network, as for expert similarity reported above, only Open > Focus ($p = .003$, ES = .64), the Full mean fell midway between Focus and Open and was not significantly different than either. Again counterintuitively, including only the central key terms in the writing prompt decreased the essay network similarity of peers to each other.

6. Scientific or scholarly significance of the study or work

This quasi-experimental investigation considered how adding central and peripheral key terms to a writing prompt influences the resulting essays. The amount of prompt directedness (i.e., Open, Focus, and Full) did not significantly influence the frequency of central and peripheral key terms in the essays. But there was a substantial difference between the essays with the Open prompt (no terms provided) compared to the essays from the Focus and Full prompts. Results show that the summary essays of those given the Open prompt *without key terms* had longer essays and the resulting networks of those essays were relatively more like the expert referent and like their peers' essays. Although tentative, these results indicate that writing prompts should NOT include key terms.

Regarding *limitations of this investigation*, because this is a quasi-experimental investigation, pre-intervention group equivalence is uncertain, and so the results here are highly tentative. However the substantive difference in essays with and without key terms in the prompt observed here warrants further research. This investigation should be replicated using an experimental design with random assignment to treatment with the addition of human-rater essay scores for concurrent validity in order to better confirm this finding regarding providing key terms.

Reason and common sense would suggest that including key terms in a writing prompt should scaffold writing and would thus result in better essays. For example, low-directed concept mapping, where no key terms are provided, tended to better reveal conceptual understanding (both explanations and errors) than did high-directed mapping that provided some or all terms and some structure (Ruiz-Primo et al. 2001).

Perhaps including key terms in a writing prompt disrupted retrieval during writing in this case because the alphabetical list of terms misaligns with both the writers' conceptual structure of this content as well as with the actual lesson conceptual structure. Research into *part-list cuing inhibition* (Slamecka, 1968) has shown that providing several items as cues counterintuitively decreases recall memory relative to no cues. Basden, Basden, and Stephens (2002) explain this inhibition by showing that cue order maters.

Extending part-list cuing inhibition from the concept map research to writing-to-learn, the lesson materials and especially the expert referent network have inherent conceptual structure, inhibition could occur when the structure of the cue (i.e., here, the alphabetical list of key terms in the prompt) mismatches the undergirding conceptual structure of the lesson. Further research is needed to determine if the key terms inhibition observed in this investigation, as well as part-list cuing inhibition, are mediated by matching or mismatching of conceptual structures of the learner, the cue, and the lesson.

This investigation was supported by the National Science Foundation, NSF award #2215807.

References

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based Writing-to-Learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29-58.

Basden, B. H., Basden, D. R., & Stephens, J. P. (2002). Part-set cuing of order information in recall tests. *Journal of Memory and Language, 47*(4), 517-529. https://doi.org/10.1016/S0749-596X(02)00016-5

Blevins-Knabe, B. (1987). Writing to learn while learning to write. *Teaching of Psychology, 14*(4), 239-241.

Clariana, R. B. (2010). Deriving group knowledge structure from semantic maps and from essays. In D. Ifenthaler, P. Pirnay-Dummer, & N.M. Seel (Eds.), *Computer-Based Diagnostics and Systematic Analysis of Knowledge* (pp. 117-130). New York, NY: Springer.

Clariana, R. B., & Koul, R. (2004). A computer-based approach for translating text into concept map-like representations. In A.J.Canas, J.D.Novak, and F.M.Gonzales, Eds., Concept maps: theory, methodology, technology, vol. 2, in the Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain, Sep 14-17, pp.131-134. See http://cmc.ihmc.us/papers/cmc2004-045.pdf.

Clariana, R. B., & Wallace, P. E. (2007). A computer-based approach for deriving and measuring individual and team knowledge structure from essay questions. *Journal of Educational Computing Research, 37* (3), 209-225.

Graham, S., & Perrin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445-476.

Lomask, M., Baron, J. B., Greig, J.,& Harrison, C. (1992, March) ConnMap: Connecticut's use of concept mapping to assess the structure of students' knowledge of science. Paper presented at the annual meeting of the National Association of Research in Science Teaching, in Cambridge, MA.

Ruiz-Primo, M. A. (2004). Examining concept maps as an assessment tool. In A.J.Canas, J.D.Novak, and F.M.Gonzales, Eds., Concept maps: theory, methodology, technology, vol. 2, in the Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain, Sep 14-17. See https://cmc.ihmc.us/Papers/cmc2004-036.pdf

Ruiz Primo, M. A., Shavelson, R. J., Li, M., & Schultz, S. E. (2001). On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *Educational Assessment, 7*(2), 99-141.

Slamecka, N. J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology, 76*, 504-513.

Visser, T., Maaswinkel, T., Coenders, F. & McKenny, S. (2018). Writing prompts help improve expression of conceptual understanding in chemistry. Journal of Chemistry Education, 95, 1331-1335

WACC (2022). *Writing across the Curriculum Clearinghouse*. Retrieved from https://wac.colostate.edu/resources/wac/intro/wtl/

Wang, Y. (2021). *Comparative Investigation of full and focused network feedback on students' knowledge structure and learning* (Unpublished doctoral dissertation). The Pennsylvania State University, State College, PA. Retrieved from https://etda.libraries.psu.edu/catalog/18696yjw5074

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1160-1175.