Effects of Document Generation and Source Presentation on Historical Understanding and Thinking

Cecil Robinson, Gina Raineri, University of Alabama, Campus Box 870231, Department of Educational Studies, College of Education, Tuscaloosa, AL, 35487-0231
Email: crobinso@bamaed.ua.edu, raine030@bama.ua.edu

Abstract: Students in an Introduction to Lifespan Development course (N = 108) participated in an experiment with a 2x2 randomized factorial design that explored the effects of document generation (control, document generation; DGT) and presentation format (single text, multiple documents) on historical understanding and thinking. As predicted, there were no improvements in students' historical understanding in the DGT or presentation format conditions. However, there were main effects for DGT and presentation format on students' depth of processing as measured by the number of added, borrowed and transformed sentences in an argument writing task. There was also a main effect for DGT on students' recall of document source information, an indication of the use of a sourcing heuristic and component of expert historical thinking. Results provide cautionary evidence for educational policy and classroom practice about document-based assessment practices, and the continued advocacy of multiple primary source document use.

A burgeoning number of cognitive studies have explored how students understand and think about historical events (Carretero & Voss, 1994; National Research Council, 2005; Stearns, Seixas & Wineburg, 2001). One prominent line of research is centered on differences between novice and expert historical thinking; that is, the fundamentally different problem-solving heuristics and epistemological frameworks between students and teachers, and historians (c.f., Wineburg, 1991; Yeager & Davis, 1995; Young & Leinhardt, 1998). This research highlights students' and teachers' practices, successes and struggles within the discipline of history, and focuses on instructional approaches that change how students and teachers think about history so that they are more closely aligned with the practices of historians (VanSledright, 2002). A second line of prominent research is focused on understanding historical events. Examples of this research include limitations of textbooks (Beck, McKeown, Sinatra, & Loxterman, 1991; Britton & Gulgoz, 1991), the effects of single and multiple documents (Wiley & Voss, 1999), and the effects of primary and secondary sources (Rouet, Britt, Mason, & Perfetti, 1996). Combined, these lines of research have provided research-supported justification for primary source document (PSD) use within classrooms that 'do history'—that is, classrooms that engage in authentic historical inquiry, and provide instructional practices as a means to foster students' historical thinking and understanding (Levstik & Barton, 1996).

During political times when educational researchers are questioning how to 'make a difference,' this research is affecting classroom practice and national policy. The National Assessment of Educational Progress (NAEP) indicates teachers are using more documents in their history classrooms (Lapp, Grigg & Tay-Lim, 2002). U.S. History standards advocate the use of primary sources (National Center for History in the Schools, 1996). An increasing number of state and national examinations are using document-based questions (DBQs) as assessment items (Grant, 2003). In 2004, federal tax appropriations of $44 million were allocated for 45 new Teaching American History grants centered on multiple documents to foster historical thinking and understanding (Department of Education, 2004).

Although this research is affecting practice and policy, there are at least two potential concerns related to multiple documents and DBQs that may limit continued adoption of these practices that need to be addressed. The first concern is that increased classroom document use has not translated into gains on achievement test scores in history. Results from the NAEP indicate that 8th and 12th grade students using multiple documents 'every day' scored *lower* than students who encountered multiple documents 'once or twice a week' (Lapp, et al., 2002). Further, there were *no meaningfully significant student performance differences* on the NAEP exam regardless if students were taught exclusively from the text or with multiple documents (Lapp, et al., 2002). These results are corroborated by research indicating that students' historical thinking is fairly resistant to change even when taught by a master teacher using well-chosen PSDs (VanSledright, 2002). These results pose a significant problem given the widespread advocacy of multiple documents, increased document use in the classroom, and allocation of federal monies to further increase document use.

The second concern is the 'authenticity' of DBQs. DBQs present six to eight documents related to a single topic—speeches, graphs, quotations, historian accounts, etc.; ask 'constructed response' questions about the main idea(s) within each document; and ask an essay response that requires students to draw on their background knowledge and from information within the documents. Advocates of DBQs argue making sense of an historical event using documents is consistent with the discipline of history and offers students an engaging way to approach learning history. They also argue that DBQs are examples of authentic large-scale performance assessments. Grant, Gradwell, and Cimbricz (2004) question claims of authenticity by arguing that presenting students with pre-selected (and often uninteresting) questions and documents without the opportunities for revisions is far removed from the work of historians, and demonstrate how these limitations raise pedagogical and assessment issues (Grant et al., 2004). They argue that with few exceptions (c.f., Wigginton, 1978) there are few 'authentic' classroom examples of students 'doing history' and concede that creating truly authentic, disciplinary-based classroom activities and assessments is generally not feasible. As an alternative, Grant et al. suggest that researchers, test-makers, and teachers attempt to create 'more authentic' activities and assessments.

More authentic classroom activities and assessments recognize that authenticity is not tenable, and instead strives to align activities and assessments with authentic historical disciplinary practices as closely as possible given classroom and/or testing constraints. For example, rather than providing students with a fixed question and predetermined set of documents, DBQs could be modified such that students: create their own significant or interesting historical questions, generate a list of potential documents to address a question and select and use documents from a large document set to address a question, to name a few. Each suggestion represents a task that is fairly easy to set up in a classroom or testing setting, and is representative of the types of tasks that historians must complete within their discipline. Although these suggestions appear to be reasonable, they have not been tested. From an instructional standpoint, research is unable to indicate whether each suggestion will significantly influence students' historical understanding and thinking. From an assessment standpoint, research is unable to indicate any of the suggestions can serve as a proxy for historical understanding and thinking.

The purpose of this study is to provide an initial examination into the effects of one of Grant et al.'s (2004) suggestions for more authentic DBQs. Specifically, this study explores whether the generation of a list of documents to address an essay question affects students' historical understanding and thinking. This study also replicates past research about single and multiple document presentation formats on understanding and thinking.

Past historical cognition research has demonstrated how manipulations to instructional tasks and assessments have significant effects on students' historical understanding and thinking. Students demonstrate significantly greater historical understanding when the phrase, "think like a historian" is added to a writing assignment (Wiley & Voss, 1996). Increases in understanding also occur when students are asked to write an argument as opposed to an essay (Wiley & Voss, 1999), a historical event is presented as a controversy (Perfetti, Britt, & Georgi, 1995), and authors' names are included or excluded on a historical text (Paxton, 1997). Beck, McKeown, Sinatra, and Loxterman (1991) demonstrated the significance of modifying the structure of the text—text coherence—on historical understanding. Multiple documents in an on-line environment produced better historical understanding than a single document, but there was no difference between multiple and single documents in paper format (Britt & Aginliskas, 2002; Wiley & Voss, 1999). Students' use of a sourcing heuristic increased when using web-based computer technology was developed to support historical thinking (Rouet, Britt, Mason, & Perfetti, 1996).

Although research has not explored whether the generation of a list of documents—a Document Generation Task (DGT)—will increase students' historical understanding and thinking, past research has demonstrated generation effects in memory tasks and higher-order mathematics problem solving (McNamara & Healy, 1995, 2000). A generation effect is produced when learners are forced to make inferences about an incomplete task. For example, memory recall of word pairs is significantly better when students are presented and asked to memorize the word pair "h_t—cold" in comparison to the word pair "hot—cold." Generating the "o" in "hot" produces semantically encoded information that results in deeper levels of cognitive processing. Deeper processing consistently leads to increases in students' ability to recall information and use previous knowledge to manipulate new knowledge for understanding (Gardiner & Hampton, 1985; Slamecka & Graf, 1978).

Given the findings from past generation effects research, one would predict that incomplete DBQ tasks that require students to generate inferences would increase students' depth of processing. One would also predict that

generating a list of documents during a DGT requires more inferences than when presented with an already established set of documents. Further, since the semantically encoded information produced during a DGT is focused on document information, and not the historical content, one would predict increases in the recall of information about the documents used to complete a DGT, but not any increases in the recall of historical content information.

Method

Participants

Students (N=108) enrolled in an Introduction to Lifespan Development class at a public research university in the southeastern U.S participated in the study. Half of the students were male (n = 54) and half were female (n = 54). 68.5% (n = 74) of the students were white, and 26.9% (n = 29) were African-American, and 76% (n = 82) of the students were freshman who had not yet established a college GPA. All students received extra credit for participation in this study.

Materials

A questionnaire was used to gather demographic information and students experiences with history instruction. A one-page summary of facts about events of the Homestead Steel Strike that occurred at the Carnegie Steel Company in Homestead, Pennsylvania in 1892 provided initial background information about the event. The *Document-Based Question* (DBQ) read:

> Although the general facts about the Homestead Steel Strike are known, the cause of the strike is not entirely known. In fact, there is still great debate and controversy among historians about who was ultimately responsible for the Homestead Steel Strike. The purpose of this study is for you to contribute to this debate and argue your opinion of events related to the strike. Your task is to think like a historian and create an argument about the controversy surrounding the Homestead Steel Strike and the extent to which Andrew Carnegie was responsible for breaking up the union. (Adapted from Rouet et al., 1996)

The *Document Generation Task* (DGT) asked students to *generate a list of materials and rationales for using those materials* that they would want to use when creating their argument. Two document sets were used in this study. The *multiple document* set consisted of a textbook excerpt, 2 historian accounts, and 4 PSDs. The *single text* contained the same information, but embedded the historian accounts and PSDs within the textbook excerpt and included transitional clauses between paragraphs (c.f., Britt & Aginlaskas, 2002). The document sets provided students with the data to construct their argument about the Homestead Steel Strike.

A multiple choice and free recall post-test was used to measure students' historical understanding, depth of processing, and historical thinking. Seven true-false questions and six free recall questions provided measures of students' understanding about the events that took place during the Homestead Steel Strike. The six free-recall items asked students to recall three supporting and three opposing arguments from the text(s) about Carnegie's role in breaking up the union. Depth of processing was measured with three situational questions (adapted from McNamara, Kintsch, Songer, & Kintsch 1996; Wiley & Voss, 1999). Each situational question had five responses: one answer item was not found within any of the readings (added), two answer items—one from the textbook and one from a document—were borrowed from the readings (surface processing), and two answer items transformed information from the text (deep processing). Six free recall questions provided measures of students' historical thinking, specifically their ability to recall information about the document sources.

Procedures

Students were randomly assigned to one of four experimental groups within a 2 (Document Generation Task, Control) x 2 (Single Text, Multiple Document) factorial design. All students completed the background questionnaire and were given the one-page summary of facts about the Homestead Steel Strike and the DBQ. A research assistant read the one-page summary and DBQ aloud to ensure that all students were exposed to the background information and the overarching task. Students in the DGT condition completed the DGT, and students in the No DGT condition were given one of the two document sets. Once the students in the DGT condition completed the DGT, they were also given one of the two document sets. After students read through the documents,

they composed their argument responses. Once they completed their arguments, they completed the historical understanding and thinking post-test assessments.

Analysis

Historical understanding was analyzed by comparing the number of true-false and free recall questions that students answered correctly on the post-test assessment across experimental conditions. Depth of processing was measured by students' responses on the post-test assessment and by decomposing students' argument responses into sentence-level propositions. The sentence-level propositions were then coded as added, borrowed or transformed (c.f. Wiley & Voss, 1999). Added sentences were sentences with information that was not found in the text. Borrowed sentences represented material that was paraphrased or quoted from the text(s). Borrowed sentences represent a surface level processing of information within the text(s). Transformed sentences were either sentences that combined two or more pieces of information from text(s) in a novel way, or combined information from the text(s) with added information from the student. In both instances, transformed sentences represent deeper levels of processing information and engagement with the text (Wiley & Voss, 1999). Historical thinking was measured by coding sentences in each written argument according to the document source(s) that were used to generate borrowed and transformed sentences, and by comparing the number of correct free-recalled source questions on the post-test measure.

Results

Historical Understanding

A 2x2 ANOVA yielded no main effects or interactions for document generation or presentation format ($p > .05$). This is consistent with our predictions about the DGT using generation effects research, and previous presentation format research and the mixed results from the NAEP.

Table 1. Measures of Historical Understanding

	Document Generation		No Document Generation	
	Single Text	Multiple Texts	Single Text	Multiple Texts
True False	3.4	3.2	3.2	3.1
Arguments	5.1	4.9	4.4	4.7

Depth of Processing

A two-way ANOVA revealed a significant main effect for presentation format on the average added, borrowed, and transformed sentences (ABT average) F (1, 84) = 7.023, p = .01, d = .54 (see Figure 1). The effect of presentation format on levels of processing is consistent with past research; multiple documents promote deeper levels of processing than single text accounts. Consistent with past generation effect research, there was a significant main effect for DGT on ABT average F (1, 84) = 6.436, p = .01, d = .57. There was no interaction between Document Type and DGT. These results indicate that the DGT promotes additional processing beyond presentation format as the single text, DGT condition produced equivalent results to the multiple text, no DGT condition. This latter finding has implications for cash-strapped classrooms. If teachers do not have access to multiple documents, the inclusion of a DGT task with a single text account can substantially increase levels of processing to levels that are similar to using multiple documents.

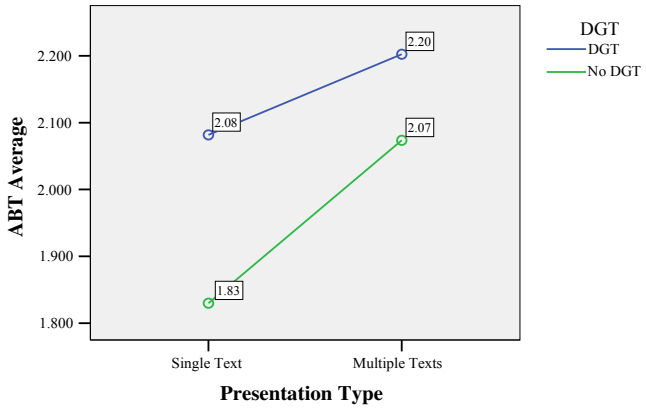

Figure 1. Level of processing by document generation task and presentation format.

Historical Thinking

A 2x2 ANOVA revealed a significant main effect for the DGT on the number of sourcing questions correctly recalled, $F(1, 107) = 4.649$, $p = 0.033$. There was no main effect for presentation format, $F(1, 107) = 0.154$, $p > 0.10$, or interaction between DGT and presentation format $F(1, 107) = 0.674$, $p > 0.10$. These results are consistent with past generation effects research, and indicate that a DGT not only leads to greater depth of processing of information, but that it also leads to greater recall of information of the information being processed. This is further verified by the non-significant increases in historical understanding.

Although there were increases in the amount of source information recalled, analysis of students' written responses indicates that DGT and presentation format did not affect the types of documents used in their argument responses. Across all conditions students used a significantly greater number of secondary documents ($M = 6.97$, $SD = 4.36$) than PSDs ($M = 1.24$, $SD = 1.67$), $t(87) = 12.10$, $p < 0.001$ in their arguments. Post hoc analysis revealed PSDs were used less than historian accounts ($M = 2.97$, $SD = 2.41$) $t(87) = 7.252$, $p < 0.001$ and textbook accounts ($M = 4.00$, $SD = 2.92$) $t(87) = 7.193$, $p < 0.001$. Textbook accounts were used more than historian accounts, $t(87) = 3.117$, $p = 0.002$. These findings are consistent with past research on historical thinking (c.f., Wineburg, 1991; Yeager & Davis, 1995) and support VanSledright's (2002) claim that students are resistant to learning from multiple documents.

Discussion

The purpose of this study was to explore the effects of presentation format and document generation on students' historical thinking and understanding. Results indicate higher student scores for the document generation condition on recall of fact-based information, but these results were not significant. This is consistent with NAEP data and warrant further exploration. If a goal of multiple document use is to increase students' historical understanding and another goal of the field is to continue to make a difference, then researchers will need to better understand the relationship between single and multiple texts and basic historical understanding.

Although there were no significant effects for recall data, there were significant effects for DGT and presentation format on the depth of processing of information within texts. The presentation format results replicate past research (Rouet et al., 1996; Wiley & Voss, 1999) and provide further evidence that suggests that multiple texts foster deeper levels of processing. The results of the DGT on depth of processing add to the historical thinking and understanding literature. These results also provide evidence that expand past generation effects research and the effects on deeper cognitive processing by demonstrating that generation effects are not only present in lower-order cognitive tasks and higher-order math tasks, but also higher-order thinking in history and writing.

From a classroom standpoint, the results of the DGT provide initial validation for Grant et al.'s (2004) argument about the authenticity of DBQs. Namely, that inclusion of 'more authentic' activities will have significant effects on the ways that students approach these tasks. However, the results also highlight the need for a detailed program of research into the differential effects of their suggestions on historical understanding and thinking. For

example, the DGT significantly affected students' depth of processing and their ability to recall source information, but did not result in significant changes in students' historical understanding or other aspects of historical thinking beyond recall of source information. Testing the other recommendations of Grant et al. beyond generating documents to include document selection and use, and question generation is a means to develop our understanding of learning from multiple documents and creating classroom environments that are not only modeled on expert disciplinary practices, but also result in the improvement of students' achievement (such as the NAEP).

In an age of accountability and high-stakes testing, any problems with or improvement to assessment items are significant. Document-based questions and PSDs are used on AP Exams, State Exams, and the NAEP assessments, and this study highlights how the addition of a simple task such as document generation creates deeper levels of processing and increased use of a sourcing heuristic. This study also replicated results from the NAEP and other experimental findings; namely that multiple documents may not significantly effect understanding any more than a well-crafted textbook. Although more research is clearly needed, this work adds caution to the wind for advocates arguing that more document be used in the classroom. This work also adds promise for the future of task effects as a means to understand how to create and understand the results of 'more authentic' historical assessments.

References

Beck, I., McKeown, M., Sinatra, G., & Loxterman, J. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly, 26*, 251-276.

Britt, M.A. & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*(4), 485-522.

Britton, B., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional texts: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83*, 329-404.

Carretero, M., & Voss, J. F. (1994). *Cognitive and instructional processes in history and the social sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Department of Education (2004). *Teaching American History Grant Program: 2004 grant abstracts*. Retrieved November 1, 2004, from http://www.ed.gov/programs/teachinghistory/ 2004tahabstracts/index.html.

Gardiner, J.M. & Hampton, J.A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 732-741.

Grant, S. G. (2003). *History lessons: Teaching, learning, and testing in U.S. high school classrooms.* Mahwah, NJ: Lawrence Erlbaum.

Grant, S.G., Gradwell, J.M., & Cimbricz, S.K. (2004). A question of authenticity: The document-based question as an assessment of the students' knowledge of history. *Journal of Curriculum and Supervision, 4,* 309-337.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Press.

Lapp, M.S., Grigg, W.S., & Tay-Lin, B.S.H. (2002). *The nation's report card: U.S. History 2001*. Washington: U.S. Department of Education.

Levstik, L.S., & Barton, K.C. (1997). *Doing history: Investigating with children in elementary and middle schools*. Mahwah, NJ: Lawrence Erlbaum Associates.

McNamara, D.S. & Healy, A.F. (1995). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory and Language, 34*, 399-416.

McNamara, D.S. & Healy, A.F. (2002). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language, 43*(4), 652-679.

McNamara, D., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.

National Center for History in the Schools (1996). *National standards for history: Basic edition*. Retrieved On-line, October 15, 2005 from the National Center for History in the Schools Web site: http://nchs.ucla.edu/standards/toc.html.

National Research Council (2005). *How students learn: History in the classroom.* A targeted report for teachers, M. Donovan and J. Bransford (Eds.), Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Paxton, R.J. (1997). "Someone with Like a Life Wrote It": The Effects of a Visible Author on High School History Students. *Journal of Educational Psychology, 89*(2), 235-250.

Perfetti, C. A., Britt, M. A., & Georgi, M. C. (1995). *Text-based learning and reasoning: Studies in history.* Hillsdale, NJ: Lawrence Erlbaum.

Robinson, C., & Cimbricz, S. (2003, November). *Assessing historical understanding: The document generation task – a glimpse of a 'more authentic' assessment task.* A paper presented at the meeting of the College and University Faculty Assembly, Chicago, IL.

Rouet, J.F., Britt, M.A., Mason, R.A., & Perfetti, C.A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88*(3), 478-493.

Slamecka, N.J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592-604.

Stearns, P. N., Seixas, P., Wineburg, S. (2000). *Knowing, teaching & learning history: National and international perspectives.* New York: New York University Press.

VanSledright, B. (2002). In search of America's past: Learning to read history in elementary school. New York: Teacher's College Press.

Wiley, J. & Voss, J.F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*(2), 1-11.

Wiley, J. & Voss, J. (1996). The effects of "playing" historian on learning in history. *Applied Cognitive Psychology, 10,* 63-72.

Wineburg, S.S. (1991). Historical problem solving: a study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*(1), 73-87.

Yeager, E.A., & Davis, O.L., Jr. (1995). Between campus and classroom: Secondary student-teachers' thinking about historical texts. *Journal of Research and Development in Education, 29(1),* 1-8.

Young, K.M., & Leinhardt, G. (1998). Writing from primary documents: A way of knowing in history. *Written Communication, 15(1),* 25-68.