SREE

2040 Sheridan Rd
Evanston, IL 60208

Society for Research on Educational Effectiveness
Advancing Education Research

www.sree.org
847.467.4001

2010 SREE Conference Abstract Template

Thank you for your interest in the Society for Research on Educational Effectiveness 2010 Annual Conference. Conference abstracts must be submitted using this template document. The template is based on the recommendations offered by Mosteller, Nave, and Miech (2004, p. 33)[1] for structured abstracts. Abstracts should follow APA Style, as specified in the Sixth Edition of the Publication Manual of the American Psychological Association.

Abstract Preparation and Submission Procedures

Save this document to your computer. Fill in each of the sections below with the relevant information about your proposed poster, paper, or symposium. Make sure to save the document again when completed. When ready, submit your abstract at http://www.sree.org/conferences/2010/submissions/

The template consists of the following sections: title page, abstract body, and appendices (references and tables and figures). Figures and tables included as part of submission should be referred to parenthetically—"(please insert figure 1 here)." The body section of your abstract should be no longer than 5 pages (single spaced, using the Times New Roman 12-point font that has been set for this document). The title page and appendices do not count toward this 5-page limit.

Insert references in appendix A of this document. Insert tables and graphics in appendix B. Do not insert them into the body of the abstract.

For questions, or for help with abstract preparation or submission, contact us at inquiries@sree.org, or 847-467-4001

[1] Mosteller, F., Nave, B., & Miech, E. (2004). Why we need a structured abstract in education research. *Educational Researcher, 33(1),* 29–34.

Abstract Title Page
Not included in page count.

Title:

Improving student learning through the use of classroom quizzes: Three years of evidence from the Columbia Middle School project.

Author(s):

Pooja K. Agarwal, Washington University in St. Louis
Henry L. Roediger, III, Washington University in St. Louis
Mark A. McDaniel, Washington University in St. Louis
Kathleen B. McDermott, Washington University in St. Louis

Abstract Submission ID: 217

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

Testing in classrooms is usually used for purposes of summative assessment, to assign student grades. Our prior research has shown that the retrieval of information that occurs during testing is a powerful enhancer of learning and retention (Butler & Roediger, 2007; Kang, McDermott, & Roediger, 2007; McDaniel, Roediger, & McDermott, 2007; for a review, see Roediger & Karpicke, 2006a). Thus, testing can serve another important purpose besides summative assessment. We have exploited this important phenomenon in our previous research, which has led to our Test-Enhanced Learning (TEL) approach.

Our proposal falls directly within the purview of the 2010 SREE Conference Theme: Research into Practice, as we have extended our prior research into practice and evaluation in a public middle school in Illinois. Moreover, many research programs are aimed at one subject (e.g., a customized biology tutorial) or a specific skill (e.g., solving algebra word problems). Our intervention, however, aspires to provide a technique or program that can be applied to many different subject matters and can invigorate learning across the curriculum and across grade levels.

The TEL approach is grounded in at least three theoretical processes that augment learning and retention: active retrieval, learning from feedback, and improvement in metacognition. First, a body of basic experimental evidence has established that active retrieval produces a powerful positive effect on later retention (Carpenter & DeLosh, 2006; Carrier & Pashler, 1992; McDaniel, Kowitz, & Dunay, 1989; McDaniel & Masson, 1985; Roediger & Karpicke, 2006b; Karpicke & Roediger, 2008). Second, experiments conducted with educationally relevant material (but not in classrooms) confirm that feedback produces significant learning gains (Butler & Roediger, 2008; McDaniel & Fisher, 1991; Pashler, Cepeda, Wixted, & Rohrer, 2005). Accordingly, a component of our TEL intervention requires that feedback be provided for all quizzes. Finally, basic research suggests that learners generally cannot judge how well they will remember previously studied information (Dunlosky & Nelson, 1994; Jang & Nelson, 2005; Koriat, 1997; Meeter & Nelson, 2003). These poor metacognitive judgments in turn negatively impact the efficacy of student-directed study activities (Thiede & Dunlosky, 1999). Theoretically, then, interventions that improve metacognition should result in more effective student-directed studying. In our TEL approach, quizzes directly provide students with information about what they know and what they do not know. Accordingly, quizzes help students identify content that is not well learned (Finn & Metcalfe, 2008) and thereby increase the effectiveness of students' self-directed (out of class) study time (Thomas & McDaniel, 2007). Quizzing may also prompt more consistent studying of target content (Roediger & Karpicke, 2006b).

Purpose / objective / research question / focus of study:

Description of what the research focused on and why.

We examined whether a test-enhanced learning program, integrated with daily classroom practices, is effective in a middle school setting. Specifically, we implemented and

experimentally evaluated a test-enhanced learning program in 6th - 8th grade Social Studies, English, Science, and Spanish classes. Although laboratory studies documenting the benefits of quizzing on learning and retention are prominent (see Roediger & Karpicke, 2006a, for an extensive review), prior to our work little experimental work has assessed the effects of quizzing in classroom settings. The absence of classroom experiments relating to the testing effect represents a critical gap in extending the basic work to educational practice. In the typical laboratory experiment, the testing effect is demonstrated for material that subjects are exposed to once and for which they have no further access for review and study. Further, even when target material is educationally relevant (e.g., a text), it is an isolated passage not related to integrated content like that representing a classes' educational objectives. By contrast, material learned in a classroom context is seen under very different circumstances. The material is typically reinforced in homework and reading assignments, it is designated as important for the students to master, and the material is part of an integrated topic domain identified as core to the curriculum.

To remedy this critical gap in verifying that the basic testing effect work can translate to effective educational practice, our ongoing work has focused on experimental evaluation of the effects of quizzing on learning course content in classroom settings. Our past three years of research at Columbia Middle School (CMS) have shown powerful positive effects of quizzing on student performance on chapter exams, semester exams, and even on final examinations given at the end of the school year.

Setting:
Description of where the research took place.

Students in Columbia Middle School (CMS) in Illinois served as participants. The school is located in Columbia, Illinois, a community about 25 minutes southeast of St. Louis. The research team has met many times with teachers, administrators of the schools (Principals, Assistant Principals), and administrators of the School District (Curriculum Coordinator, District Superintendent). Columbia Middle School (CMS) enrolls students in grades 5-8, with a total enrollment of about 530 students. During the past three years, we have created a positive, enthusiastic, and cooperative atmosphere with CMS students, teachers, administrators, and parents.

Population / Participants / Subjects:
Description of participants in the study: who (or what) how many, key features (or characteristics).

Approximately 400 6th, 7th, and 8th grade middle school students, including special education and gifted students, participated in this research. Students in CMS are about half male and half female. Ninety-seven percent of students are Caucasian. The principal of the nearby high school (in the same school district) estimates that 75% of the graduating seniors go on to some form of further education (counting community colleges and technical trade schools).

Intervention / Program / Practice:
Description of the intervention, program or practice, including details of administration and duration.

We used chapter material drawn from the assigned textbook for each subject (Social Studies, English, Science, and Spanish). On initial classroom quizzes (pre-tests before the teacher's lesson, post-tests after the teacher's lesson, and review tests a few days later), half of the target facts from each chapter were tested in a multiple-choice format (tested condition) and half of the facts were not tested (non-tested condition), following a within-subjects design. Target facts were randomly assigned to the two conditions and each of the six classroom sections received a different random selection. The number of target facts varied between conditions and chapters.

For example, a multiple-choice fact included:

What is Pharaoh Tutankhamun best known for?

a) The way he ruled his kingdom
b) Living to an old age
c) The belongings found in his tomb
d) His trading routes with other kingdoms

For initial quizzes (pre-, post-, and review), an experimenter administered the classroom quizzes orally and visually using a clicker response system (Ward, 2007). Students were provided with immediate feedback in the form of a green checkmark next to the correct answer while the experimenter read aloud the question stem and correct answer. Questions on the initial quizzes were presented in the order in which they appeared in the chapter. The four multiple-choice alternatives were presented in a different random order for each pre-, post-, and review test.

Subjects were tested in classroom sections ranging from 21 to 27 students each. Before the teacher's lesson, students took a pre-test over tested items. The teacher was not present for the pre-test and did not know which target facts were tested or non-tested. Following the pre-test, the teacher taught the lesson for the day, which covered all target facts, both tested and non-tested facts. Immediately after the lesson, students took a post-test over tested items. Approximately two days later, students took a review test over tested items.

Research Design:

Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

We used a true experimental design, in which the manipulated TEL intervention occurred within-student, such that some materials received normal classroom exposure and other materials were assigned to the treatment condition (additional quizzing), with materials counterbalanced across students. This within-students design feature provides several advantages to the more common between-classroom, between–students design. First, power is maximized. The classroom experiments conducted in our project had extremely high power to detect a .10 effect (a small size effect): power = .995 (alpha=.05, two-tailed). Second, the within-students design precludes the potential ethical issue associated with designs in which some students have potential benefits in course performance (because of the testing intervention) and other students shoulder the costs of being deprived of the testing intervention (no-test control). Indeed, the Columbia school administrators raised this concern during our initial contacts with them, stimulating our implementation of within-subject manipulations.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

To measure retention, the classroom teacher administered chapter exams in paper-and-pencil format. Students completed a multiple-choice test comprised of all tested and non-tested items. Multiple-choice questions on the chapter exams were the same as those on the initial classroom quizzes, presented in a different random order for each classroom section. The four multiple-choice alternatives were also reordered randomly. Students received delayed feedback from the classroom teacher approximately 2 days after the chapter exam.

Students also completed multiple-choice end-of-the-semester and end-of-the-year exams, which were administered via the clicker response system to aid in data collection. All facts were tested at least once on the chapter exam, yet items on the end-of-the-year exam were not presented on the end-of-the-semester exam. Questions were presented in the order in which the chapters appeared in the textbook and questions for each chapter were presented in a different random order for each classroom section. For example, items from chapter 4 were presented in random order followed by items from chapter 5 presented in random order, etc.

Students who declined to participate, students in special education or gifted programs, and students who were not present for all initial quizzes, final exams, and delayed exams were excluded from our data analyses. Initial and final test performance was analyzed using repeated measures of analysis of variance (ANOVA). Planned t-tests were also used in order to determine significant differences between specific conditions.

Findings / Results:
Description of main findings with specific details.

Our evidence to date indicates that TEL greatly enhances student learning in courses in the middle school curriculum. Across different materials, students (including special education and gifted students), class schedules, subject matter, teachers, and classrooms, a consistent test-enhanced learning effect was obtained: students better remember information that was previously tested using classroom quizzes, in comparison to information that was re-read or not tested. This effect was shown to persist over lengthy retention intervals, even up to a year after initial classroom testing.

For example, on chapter exams in 8th grade Science (please insert Figure 1 here), the TEL program has typically taken student performance from the range of 75% baseline performance on chapter tests to around 90% or slightly better after our intervention. Teachers report to us that the baselines in our experiments are about where performance usually is for a class, so the gain is impressive. Given that performance begins at 75%, we are able to take students from roughly a C+ grade to an A- grade in the typical grading distribution used by the school. In the chapter test results below, the TEL program boosted grades by a proportional score of 65% (that is, [(.91-.74)/(1.00-.74)]×100)]. On semester exams, a significant testing effect was obtained such that student performance was still 7% greater on tested items than on non-tested items. Even at the end of the school year, a significant 7% testing effect was demonstrated for material studied during the fall semester that year. (These data indicate 29% and 17% proportional gains, respectively, above baseline performance). The data in the figure below are representative of many experiments and reveal the power and robustness of test-enhanced learning.

Conclusions:
Description of conclusions and recommendations based on findings and overall study.

A test-enhanced learning program can be successfully implemented in a classroom setting. TEL works well in subjects that are heavily fact-based (social studies, history, science, some aspects of mathematics) and in learning vocabulary (either in English or in foreign languages). In such courses, students are responsible for learning a wealth of facts pertaining to the subject matter or a fundamental set of vocabulary terms. Of course, higher order thinking skills (reasoning, solving problems, and creatively transferring knowledge to new domains) are also critical parts of the educational process, but unless students have mastered the basic structure of knowledge within a domain, they have no hope of creative applications of their knowledge (e.g., see Willingham, 2009). We have also collected evidence showing that quizzing improves student metacognition (knowing what they know and do not know) and transfer of learning (application of the knowledge to new situations).

The educational implications of this research extend to curriculum and teaching practices: When facilitating long-term learning, educators and students should be encouraged to use quizzes as a method to enhance learning.

Appendices
Not included in page count.

Appendix A. References
References are to be in APA version 6 format.

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*, 514-527.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*, 604-616.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative processing explanation of the testing effect. *Memory & Cognition, 34,* 268-276.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition*, 20, 632-642.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*, 545-565.

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. J*ournal of Memory and Language, 58*, 19-34.

Jang, Y., & Nelson, T. O. (2005). How many dimensions underlie judgments of learning and recall? Evidence from state-trace methodology. *Journal of Experimental Psychology: General*, *134*, 308-326.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528-558.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966-968.

Koriat, A. (1997). Monitoring one's knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349-370.

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, *16*, 192-201.

McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition*, *17*, 423-434.

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 371-385.

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14,* 200-206.

Meeter, M., & Nelson, T. O. (2003). Multiple study trials and judgments of learning. *Acta Psychologia*, *113*, 123-132.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 3-8.

Roediger, H. L. & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210.

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255.

Thiede, K. T., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1024-1037.

Thomas, A. K., & McDaniel, M. A. (2007). Metacomprehension for educationally relevant materials: Dramatic effects of encoding–retrieval interactions. *Psychonomic Bulletin & Review*, *14*, 212-218.

Ward, D. (2007). eInstruction: Classroom Performance System [computer software]. Texas: eInstruction Corporation.

Willingham, D. T. (2009). *Why students don't like school*. San Francisco: Jossey-Bass.

Appendix B. Tables and Figures
Not included in page count.

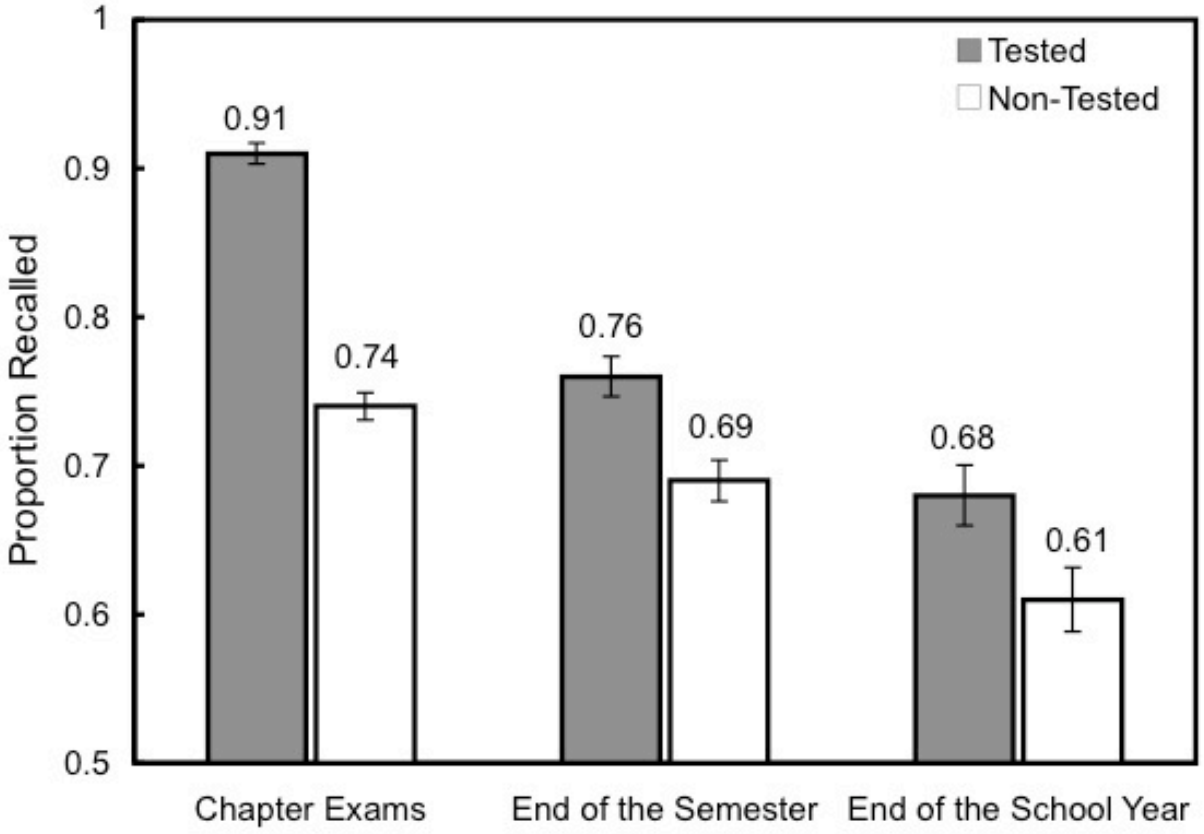

Figure 1. Student retention (proportion recalled) on chapter, end of the semester, and end of the school year exams, for tested (quizzed) vs. non-tested items in 8th grade Science.