

College Students' Memory for Unannounced Cumulative Items on the Final Exam

Lola Aagaard, Jenny Templeton, Timothy W. Conner II, and Ronald L. Skidmore

Morehead State University

Paper presented at the annual meeting of Midsouth Educational Research Association,
Knoxville, TN

November, 2014

College Students' Memory for Unannounced Cumulative Items on the Final Exam

Research regarding methods and manipulations that serve to enhance the retention and recollection of information that has been presented in various learning contexts has been a mainstay in the area of teaching and learning for some time (e.g., Ausubel, 1966; Thorndike, 1912). Of particular interest are the areas of distributed practice and the testing effect.

Research has shown the benefits of distributed practice (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006), in which the encoding of a given amount of information is undertaken during the course of several incremental study sessions rather than during a single session (i.e., massed practice or "cramming"). Distributed or distributive practice is considered to be an effective study strategy, enhancing information retention and subsequent retrieval (Rohrer & Taylor, 2006).

Testing effect refers to the phenomenon that occurs as a result of retrieval practice when recollection of information by the learner during a testing condition enhances long-term retention and recollection of that information in subsequent testing conditions. In an experimental study with college students, Karpicke and Roediger (2008) found that repeated presentation and retrieval of the same items on subsequent tests improved long-term retention of the items to a greater degree than did repeated study sessions (i.e., encoding of information without testing). McDaniel, Agarwal, Huelser, McDermott, & Roediger (2011), found that review quizzing resulted in significant performance gains on cumulative tests and end-of-year exams of middle school students. A review of studies appearing in the literature between 2006 and 2012 was conducted by Eisenkraemer, Jaeger, and Stein (2013), and more broadly presents the benefits associated with the testing effect.

Similar to the conditions presented above regarding the testing effect, cumulative tests are typically constructed by combining material being presented in the current unit of study with

that which has been presented in prior units. Although considered a bane by students, the few studies that are available do indicate benefits to cumulative testing (Lawrence, 2014; Rohm, Sparzo, & Bennett, 1986; Szpunar, McDermott, & Roediger, 2007) similar to those associated with the testing effect. As baseline information for a future study of cumulative testing, the current study investigated how well college students recalled previous material when given unannounced cumulative items on a final exam.

Method

Four sections of an undergraduate human development class formed the sample for this research, with a total sample size of 154. Two professors taught two sections each, with student demographics as shown in Table 1.

Table 1

Student Demographics per Professor

Professor	n	% Male	% Freshman	% Sophomore	% Junior	% Senior
A	76	30	17	55	24	3
B	78	28	60	29	6	0

As Table 1 shows, the students in the sections taught by Professor B were substantially younger than those in Professor A's sections.

The professors each administered three separately-developed multiple-choice exams throughout the semester. Professor A held face-to-face review sessions for students before each exam, while Professor B made a study guide available to students a week prior to each exam. For the final exam, each professor chose ten items from each of the first and second tests of the semester to include on the final exam. Students were not alerted to the presence of 20 cumulative items on the final until the exams were distributed, at which time they were asked to

do their best on those items, although the score for them would not be included in their exam grade.

Using a paired t-test, the number of students missing the items on the final was compared to the number missing the same items earlier in the semester, with the unit of analysis being exam item (N=40; 20 items per professor). A Pearson correlation was also calculated to see whether there was a relationship between the number of students missing items on the first tests vs. the final exam. These tests were conducted for the overall sample, as well as by professor.

Results

The paired t-test (see Table 2) showed a statistically significant increase in the average overall number of students missing the items on the final exam, which was expected. The size of the difference did not depend on whether the cumulative items were originally from Test 1 or Test 2.

Table 2

Descriptive and Statistical Results for Number of Students Missing Items

<u>Test</u>	<u>n</u>	<u># Students Missing Items</u>		<u>t</u>	<u>df</u>	<u>p</u>
		<u>Mean</u>	<u>Std.</u>			
<u>Overall</u>						
Test 1 or 2	40	10.1	4.83	-4.939	39	<0.001
Final	40	15.2	7.44			
<u>Test 1 items only</u>						
Test 1	20	10.1	4.98			
Final	20	14.9	6.51			
<u>Test 2 items only</u>						
Test 2	20	10.2	4.81			
Final	20	15.6	8.43			

Looking at results separately for each professor, however, did show an unexpected difference. A larger average number of students missed items on the final in Professor B's sections as compared to Professor A (see Table 3).

Table 3

Number of Students Missing Items, Separated by Professor

<u>Test</u>	<u>n</u>	<u># Students Missing Items</u>	
		<u>Mean</u>	<u>Std.</u>
<u>Professor A, Overall</u>			
Test 1 or 2	20	9.9	6.27
Final	20	12.3	6.51
<u>Professor B, Overall</u>			
Test 1 or 2	20	10.4	2.91
Final	20	18.1	7.32

The effect size of the difference between the first testing and the final testing for average number of students missing these 20 items was 0.38 for Professor A and 1.15 for Professor B.

A Pearson's correlation between the overall number of students missing each item before and on the final was 0.51, indicating that items previously missed by larger numbers of students were not necessarily the ones missed by the most students on the final exam. This correlation differed by professor, as well (0.65 for Professor A; 0.45 for Professor B).

Discussion

Generally speaking, the results do support the hypothesis that more students would be likely to miss the cumulative exam items on the final exam than they would on the earlier exams. In both classes, more students missed the cumulative items on the final than those same items when provided in Exam I or Exam II. This finding could be for several practical reasons. It is understandable that students might not recall answers to exam items related to content covered

earlier in the semester without some type of maintenance rehearsal or memory strategies being employed or encoding the material as “survival-relevant” (Miller, 2011, p. 120). Since students were not alerted to the cumulative items until the items were presented on the final exam, students would not have had foreknowledge regarding what content from the first and second exam would need to be studied for successfully answering the cumulative items.

Problems associated with testing anxiety may also have contributed to the differences noted on number of students that missed items between earlier exams and the final. This reason is somewhat unlikely as students were told that the cumulative items would not count as part of their final exam grade; however, test anxiety has been widely documented (Davis, DiStefano, & Schutz, 2008) and might have been triggered from the presence of previously unannounced content in a high stress context (the end of an academic term).

Although the hypothesis was supported generally, the differences in number of students missing the cumulative items on the final exam compared to when presented on previous exams varied between Professor A and B, suggesting that external factors may contribute to whether or not the stated hypothesis is more or less likely to occur. First, the professors did not use common tests, so there may have been differences in the test items that accounted for the number of students missing them in the different course sections. Secondly, the students in Professor A's sections of the course were provided with face-to-face review sessions prior to the administration of course exams while Professor B's students were provided with written study guides for each exam. Professor A's students, if present during the reviews, would have discussed the content out-loud with the professor. Additionally, the students would have been able to ask questions aloud, presumably followed by answers that all the students in the classes benefitted from hearing.

However, this oral review did not occur for the students in Professor B's classes. Perhaps students are less inclined to initiate studying for an exam from guides that are written, rather than verbally reviewed. Further, students in Professor B's classes might have been less likely to seek clarification on review material they did not understand as doing so would have required a separate action (asking after a class session or sending an email, for example). This difference between Professor A and B in review format may be a variable that contributed to the noted differences between the studied student groups.

In addition to the difference in review format for exams by professor, demographic data show that students in Professor B's courses were, on the whole, much younger than the students in Professor A's courses. For example, Professor A's sections only had a combined total of 17% of students in the Freshmen class whereas Professor B's sections had a much larger percentage of Freshman students at 60%. Younger students that are likely less experienced at engaging in appropriate studying strategies might reasonably be expected to miss more items on exams than older, more experienced undergraduates who may be more self-regulated, as self-regulation is related to test performance (Kitsantas, 2002; Sundre & Kitsantas, 2004).

Further, students with more experience with engaging in the learning of college-level material are also more likely, due to increased exposure to college-level rigor and prior related content from other courses in their field of study, to comprehend post-secondary level content than would be the less experienced students (like those students found in larger quantities in Professor B's class sections). This difference may also be influenced by differences that age and experience provide in the students' abilities to successfully engage in metacognitive tasks, which have been shown to be related to being able to learn (Kornell & Metcalfe, 2006).

The much older student sample in Professor A's courses might have come equipped with higher metacognitive behavioral practices and stronger self-regulatory behaviors than the much

younger students in Professor B's sections. Logically then, one result from these potential differences in academic characteristics between the students in Professor A and Professor B's class, might be the differences noted in number of cumulative items missed on the final exam.

Limitations

There are several limitations of this study that should be considered when interpreting the findings. First, although the hypothesis that students would miss more of the items from Exams I and II when presented again on the final exam was supported, the data are of students that were chosen in courses that were convenient to the researchers. Future studies would benefit from using students enrolled in the same course (an undergraduate developmental psychology for educators class), but with randomly chosen campus sites. Professors of these courses could randomly be assigned differences in how students are prepared for exams, resulting in a stronger research method that could better support conclusions made about the study hypothesis across geographical and institutional lines.

Not only were the data from students that happened to be enrolled in courses the professors were teaching, but also, the information collected about the students was not sufficient for a robust comparison of missed test items among different student characteristics. A weakness that resulted in this study was in the lack of availability of additional student characteristics data that could have served as items of analysis in the results and conclusions of the research. Individual demographic information (gender, age in years, etc.) and additional student academic characteristics (reported study habits, reported preparation for each of the exams, overall g.p.a., etc.) would have allowed for a more contextualized reporting of the results, and perhaps a stronger understanding of the relationships that are currently included in this research. In the future, care should be made in choosing what information about the subjects should be included

in the study such that data interpretation will be significant and provide meaningful interpretations of the results.

Additionally, and more importantly, another limitation to this study lies in the fact that the results are based upon collection of exam score data for students over one academic term only. In future studies, researchers could improve the study through examining data across students enrolled in this particular course over time. Also, researchers would be equipped with larger data sets from which additional inferential analyses could be utilized with the purpose of better understanding relationships between the variables examined in this study.

Finally, using data from student exam performances in a particular undergraduate course might not generalize to other subjects or courses in fields of study outside of an undergraduate developmental psychology for educators class. In order to better understand the relationships between and among exam review formats, the number of students that miss cumulative exam items from earlier exams that appear on an end-of-course final exam, and any other relationships that are appropriate to this vein of study, researchers really should consider the inclusion of student groups in common courses that are of differing subject matter, reducing the likelihood that results are informed by the subject matter being assessed (a variable not studied here). If researchers could study missed exam items in more than one content area, such a study could yield results and conclusions that are more valid than those provided from the study of students in one content area only.

As can be seen, there is still much work to be done in terms of understanding factors related to student performances on tests generally and students' likelihood of missing cumulative exam items on a final exam in college, specifically. Whether or not future researchers expand the length of time and sections of study, the content areas in which assessment data are gathered, the places where the courses are taught, the student characteristic information, or any combination of

these remains to be seen. That said, any and all of those future studies should provide useful information that will help those interested better understand the factors related to student performance on exams and help educators that are concerned with student performance better ascertain if the results in this study are in fact, valid.

References

- Ausubel, D. P. (1966). Early versus delayed review in meaningful learning. *Psychology in the Schools, 3*, 195–198.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Roher, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354-380. doi: 10.1037/0033-2909.132.2.354
- Davis, H. A., DiStefano, C., & Schutz, P. A. (2008). Identifying patterns of appraising tests in first-year college students: Implications for anxiety and emotion regulation during test taking. *Journal of Educational Psychology, 100*(4), 942-960. doi: 10.1037/a0013096
- Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A systematic review of the testing effect in learning. *Paideia (Ribeirão Preto), 23*(56). Retrieved June 25, 2014, from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-863X2013000300397&lng=en&nrm=iso
- Karpicke, J.D. & Roediger, H.L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966-968. doi: 10.1080/09541440601056620
- Kitsantas, A. (2002). Test preparation and performance: A self-regulatory analysis. *Journal of Experimental Education, 70*(2), 101-113. Retrieved from EBSCOhost ERIC database.
- Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 609-622. Retrieved from EBSCOhost ERIC database.
- Lawrence, N. K. (2014). The benefits of cumulative exams. In *Proceedings of the Conference on Higher Education Pedagogy* (pp. 168-169), Blacksburg, VA, February 5-7. Retrieved from <http://www.cider.vt.edu/conference/proceedings/2014ConferenceProceedings.pdf>

- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L.III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology, 103*(2), 399-414. doi: 10.1037/a0021782
- Miller, M. D. (2011). What college teachers should know about memory: A perspective from cognitive psychology. *College Teaching, 59*, 117-122.
- Rohm, R. A., Sparzo, F. J., & Bennett, C. M. (1986). College student performance under repeated testing and cumulative testing conditions: Report on five studies. *Journal of Educational Research, 80*(2), 99-104. Retrieved from EBSCOhost Education Full Text (H. W. Wilson) database.
- Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*(9), 1209–1224.
- Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology, 29*(1), 6-26. doi: 10.1016/S0361-476X(02)00063-2
- Szpunar, K. K., McDermott, K. B., & Roediger III, H. L. (2007). Expectation of a final cumulative test enhances long-term retention. *Memory & Cognition, 35*, 1007-1013. Retrieved from EBSCO Electronic Journal Service.
- Thorndike, E. L. (1912). The curve of work. *Psychological Review, 19*, 165–194.