Abstract

Graduate-level education in the health sciences builds from and is dependent upon knowledge supplied by prerequisite coursework. Such previously learned information may be stored well in memory without being readily accessible for application to facilitate new learning. Stored, but not retrievable information is termed "marginal knowledge"; psychology laboratory research has demonstrated that multiple-choice questions can be used to stabilize access to or "reactivate" marginal knowledge, even without feedback. Our study assessed whether this phenomenon was reproducible in an authentic classroom setting.

One hundred and forty-five student pharmacists enrolled in the first-year Pharmacy Bridging Course (PBC) participated in this study, first completing a pre-course assessment comprising fill-in-the-blank questions about material covered in undergraduate prerequisites, but not to be covered in PBC. Students were then randomized to one of two groups to study three of the six lowest performance topics per the pre-assessment. After one week of PBC, students completed an interim assessment including study questions provided as fill-in-the-blank and then in multiple-choice format, without feedback. At the completion of PBC two weeks later, students repeated the pre-course assessment as a final exam.

The interim multiple-choice test conferred better final exam performance on questions that students did not answer correctly on the pre-test, relative to the control condition (i.e., not taking the interim test; d = 1.03). In addition, the benefit of an interim multiple-choice question was significantly greater when students answered it correctly relative to incorrectly (d = 0.54). Interestingly, even incorrect responses on the interim test conferred a performance advantage over not taking the interim test (d = 0.61). Our study provides initial data to support that a multiple-choice test (even without feedback) in a genuine health professional classroom can produce a sufficient stimulus to stabilize marginal knowledge as demonstrated by improved retrieval of previously assessed material. We also know that this effect persists for at least two weeks as evidenced by the final exam performance; tracking this effect over longer follow-up periods provides an opportunity for further investigation.

Background & Purpose

Theories of human memory and learning have long drawn a distinction between the availability and accessibility of knowledge in memory (Tulving & Pearlstone, 1966). Availability typically refers to whether or not information has been stored in memory, while accessibility refers to the ability to retrieve the information at a given moment. A common example that illustrates this distinction is the "tip-of-the-tongue state," in which a person cannot retrieve a particular word or phrase, yet has a strong feeling that he or she knows it (Brown & O'Neil, 1966). Often, such states are resolved during the experience or shortly thereafter, once the

word or phrase becomes accessible (Brown, 1991). Numerous theories explicitly incorporate this distinction as part of their framework; for example, the relatively new theory of disuse (Bjork & Bjork, 1992) distinguishes between storage strength and retrieval strength, which are indices of availability and accessibility, respectively.

The investigation into stabilizing access to marginal knowledge is in its infancy, with only basic research studies completed to date, all of which in an artificial (laboratory) setting. This expanding body of research indicates that there are, in fact, several ways to stabilize access. One method shown to be effective is to re-expose subjects to the information so that they can re-learn it (Berger et al, 1999; Cantor et al., 2015). This method corresponds to the traditional educational practice of re-teaching the material. Perhaps more interestingly, researchers have also examined whether the act of taking a test can re-activate marginal knowledge. Both an open-ended test with feedback (Berger et al, 1999) and a multiple-choice test with or without feedback (Cantor et al., 2015) have been shown to be effective in stabilizing access to marginal knowledge.

Given the educational implications of this finding and the fact that its generalizability has yet be established, the present study examined whether marginal knowledge could be stabilized with a testing setup in an authentic classroom setting. Of the various methods that have been proven to be effective, taking a multiple-choice test without feedback was chosen for three primary reasons. First, it provides the most compelling evidence for the stabilization of marginal knowledge in that the other methods leave open the possibility that new learning has occurred (though laboratory studies have included additional controls to rule out this possibility). Second, in contrast with passively re-studying material, taking a test actively engages learners and provides formative information to guide future teaching and individual learning. Third, multiple-choice tests provide a more objective assessment than open-ended tests; this enables easier grading and compilation/evaluation of study data, especially where technology (e.g., exam software) is used.

Throughout a student's educational career, an expansive amount of knowledge is acquired and developed. Over time, knowledge that was highly accessible within close proximity of the learning process becomes inaccessible over time, typically at a rate commensurate with its disuse (i.e., knowledge that is not used becomes inaccessible more rapidly; Bahrick & Hall, 1991). Knowledge that has been stored in memory but cannot be retrieved is termed "marginal knowledge"; from an educational standpoint, this phenomenon is problematic in that new learning often necessitates building upon previously obtained knowledge, and can therefore be impeded if prior knowledge is not readily accessible (Berger, Hall, & Bahrick, 1999). As such, the irretrievability of marginal knowledge that accrues over time is a veritable concern that merits critical assessment and evaluation across all levels of learning, from the primary (e.g., loss of knowledge during a winter or summer break) to the secondary and graduate educational spheres (Cooper et al., 1996).

A common approach to the marginal knowledge problem involves implementation of a designated "remediation" or "review" component of a course, typically an initial block of time devoted to covering material that students should already know from prerequisites. Some curricula employ formal "bridging" courses as a tool to facilitate a smooth transition between schools or programs, creating a safety net to ensure that requisite knowledge is present to enable new learning; however, the use of a bridging course is a sub-optimal solution as it necessitates the use of classroom time that could otherwise be dedicated to instruction on novel topics (i.e., facilitating new learning). The goal of our study was to investigate a potential alternative solution to this problem. Because laboratory research has suggested that access to marginal knowledge can be stabilized through exposure to a multiple-choice test (Cantor et al, 2015), we investigated whether taking a multiple-choice test can stabilize access to marginal knowledge in an authentic, health professional classroom setting.

Study Design & Methods

One-hundred and forty-five first-year pharmacy students participated in the study (95% consenting); these students were enrolled in a bridging course designed to help them transition into the Doctor of Pharmacy program at the University of North Carolina Eshelman School of Pharmacy. The average age of the students was 22 years old (range 19-32) and 81% of them had obtained a baccalaureate degree. All students were required to have completed at least one semester of human anatomy and physiology prior to entering pharmacy school (i.e, this was a prerequisite subject). As a group, their mean college grade point average (GPA) upon admission was 3.5 (out of 4.0) and their mean Pharmacy College Admission Test (PCAT) score was 88%. IRB approval was obtained (Study # 15-0783).

This study was conducted as part of standard educational activities within the "Pharmacy Bridging Course" (PHCY 500) for first-year PharmD students. This four-week course takes place at the beginning of the fall semester, and reviews core material from prerequisite coursework in five areas: physiology, organic chemistry, biochemistry, applied math, and statistics. All five subject areas are reviewed in parallel during the bridging course, which is designed to maximize the likelihood of student success in the curriculum with the following goals: (1) to give students a chance to develop behavior patterns conducive to academic success in the curriculum; (2) to review key facts, principles, and concepts from relevant science and math disciplines with a pharmacy perspective; (3) to level the playing field for a diverse group of students so that each can excel in subsequent foundational courses; and (4) to introduce students to the instructional methods that they will encounter across their foundational courses. For each subject area within PHCY 500, a pre-course assessment is administered to measure proficiency based on prior coursework and experiences; a post-course assessment is also required to assess content mastery following completion of the course.

Students completed a pre-course physiology assessment that consisted of 93 questions about material that they should have learned in prior courses per undergraduate syllabi review. The assessment was administered using ExamSoft software (ExamSoft, ExamSoft Worldwide, Inc. © 2012). Approximately 70% of this pre-course assessment was about material that would be reviewed in the course ("current course material"), whereas the other 30% of the assessment was about material that would not reviewed in the course but was important for later coursework in the program ("future course material"). More specifically, students answered 65 multiple-choice questions about current course material that covered seven areas of physiology (cardiology, renal, hematology, central nervous system, gastrointestinal, respiratory, and endocrine). They answered 28 fill-in-the-blank questions about future course material with four questions from each of the seven topics.

The experimental condition focused solely on the future course material (i.e., material that was previously learned in prerequisites and will be important for future learning) because it would not be covered during the PHCY 500 bridging course. Following completion of the pre-course assessment, the seven topics were ranked from highest to lowest based on overall student performance on the future course material. The topic on which students performed the best was excluded from the experiment, with the rationale being that this topic area would have the least amount of marginal knowledge available for reactivation if the students were fairly comfortable with the subject at the time of the pre-course assessment. The six remaining topics were included in the experiment; these were divided into two approximately matched sets. A single independent variable (interim multiple-choice test condition: test, no test) was manipulated within-subjects and between-materials. That is, students were randomly assigned to receive a multiple-choice test during the first week of the course on one set of topics about the future course material (i.e., either Topics 2, 4, and 6 or Topics 3, 5, and 7). The assignment of each set to experimental condition was counterbalanced across participants.

During the first week of the PHCY 500 course, students took an interim test on the assigned set of three topics using OpenStax Tutor, a computerized, personalized learning system (OpenStax Tutor, Rice University © 2015). The interim test consisted of the same four questions for each topic that had appeared on the pre-course assessment; however, the format for responding was different. First, students were required to respond to the question in fill-in-the-blank format. After they had submitted a response, they were provided with the same question in multiple-choice format, and asked to select the correct response from among the alternatives. The interim test questions about the future course material were given to students as part of a larger set of practice questions about the current course material from the pre-course assessment; that is, questions about the future course material were intermixed with current course material within the same practice assignments. Although no feedback was provided for the questions about future course material, students did receive feedback for the questions about current course material.

Students were responsible for completing all practice for the three assigned topics during the first week to ensure that they would compete the interim test questions for half of the future course material. They were also required to complete the practice for the other four topics, but this practice only included questions about the current course material and it could be completed at any time before the end of the third week of the course. Grading was based only on completion of the practice by the assigned deadline (i.e., not on performance). In addition to enabling the experiment, distributing practice on the topics had the practical benefit of helping students with time management and reducing the overall amount of work each week during a time compressed course. During the fourth and final week of the course, students completed the post-course assessment, which was identical to the pre-course assessment, with the exception that after taking the assessment, they received feedback on all of the questions (i.e., both the current and future course material) to ensure that they had the correct information moving forward to prepare for subsequent coursework.

Data Analysis & Results

Paired sample t-tests were performed using SPSS 23 software (IBM, 2016). Cohen's d was selected as the parameter for measuring effect size; this is a popular statistic used in psychology research. The interpretation of the Cohen's d value is not particularly straightforward without an understanding of the data's context and standard deviation, but a general interpretation is that a Cohen's d value of 0.2 indicates a small effect size, a value of 0.5 indicates a moderate effect size, and a value of 0.8 indicates a large effect size when interpreting the effect of an intervention (Cohen, 1977). Statistics are reported in APA format. As stated above, the results reported only pertain to the future course material.

Pre-Course Assessment: As expected given that the testing (experimental condition) had yet to occur, there was no difference in the proportion of correct responses on the pre-course assessment between the items in the topics assigned to the interim test vs. no interim test conditions [M = 0.24 vs 0.24; $t < 1$].

Interim Multiple-Choice Test*:* Students correctly answered approximately two-thirds of the questions on the interim multiple-choice test (M = 0.65). It was rare for a student to get a question correct on the pre-course assessment and subsequently answer it incorrectly on the interim test (Odds Ratio: 0.06 [0.042, 0.096]).

Post-Course Assessment. Figure 1 depicts the main results for the post-course assessment. For items that students did not answer correctly on the pre-course assessment, taking the interim multiple-choice test led to significantly better performance on the post-course assessment relative to not taking the interim test [M = 0.42 vs 0.17; $t(144) = 10.64$, SEM = 0.02, $p < 0.001$, d = 0.88]. There was also a minor, but statistically significant benefit of the multiple-choice test for items that were correctly recalled on the pretest relative to the control condition [0.88 vs. 0.81;

t(137) = 3.41, *SEM* = 0.02, *p* = 0.001, d = 0.27]. Within the interim test condition, the benefit was significantly greater when participants succeeded in answering the multiple-choice question correctly relative to answering it incorrectly [*M* = 0.50 vs. 0.33; *t*(141) = 6.79, *SEM* = 0.03, *p* < 0.001, d = 0.57]. Interestingly, even when participants answered an item *incorrectly* on the interim test, taking the test conferred an advantage over not taking it [*M* = 0.33 vs. 0.17; *t*(142) = 5.39, *SEM* = 0.03, *p* < 0.001, d = 0.46].

Discussion & Significance

Performance on the post-course assessment evinced that a single exposure to a multiple-choice question, even without feedback, was sufficient to stabilize access to marginal knowledge (i.e., previously learned knowledge that was inaccessible before the bridging course); this effect persisted for two weeks. Two notable limitations of this study are (1) there was no definitive way to determine that students had previously learned content tested by the marginal knowledge study questions, which may have minimized the effect of the testing condition if students had not in fact learned this material previously and had it available for reactivation; and (2) there was a minor amount of content overlap within the other subjects of the bridging course (e.g. biochemistry, organic chemistry) that may have provided students with re-instruction on topics covered within the study questions, which may have positively skewed the effect of the testing condition. Even so, our study demonstrates the generalizability of previous laboratory findings in this area to an actual classroom, thus opening a door into opportunities for future research using this approach to improve educational practice. A particularly promising application of this finding might be the development of strategic interventions to re-activate relevant marginal knowledge, timed to occur prior to new learning. Given recent advances in educational technology, it will be possible to identify and activate marginal knowledge using testing methods at an optimal time. A potential direction for future research is to examine whether such strategic re-activation of marginal knowledge can persist for longer periods of follow-up, and may facilitate *new* learning in a genuine classroom setting.

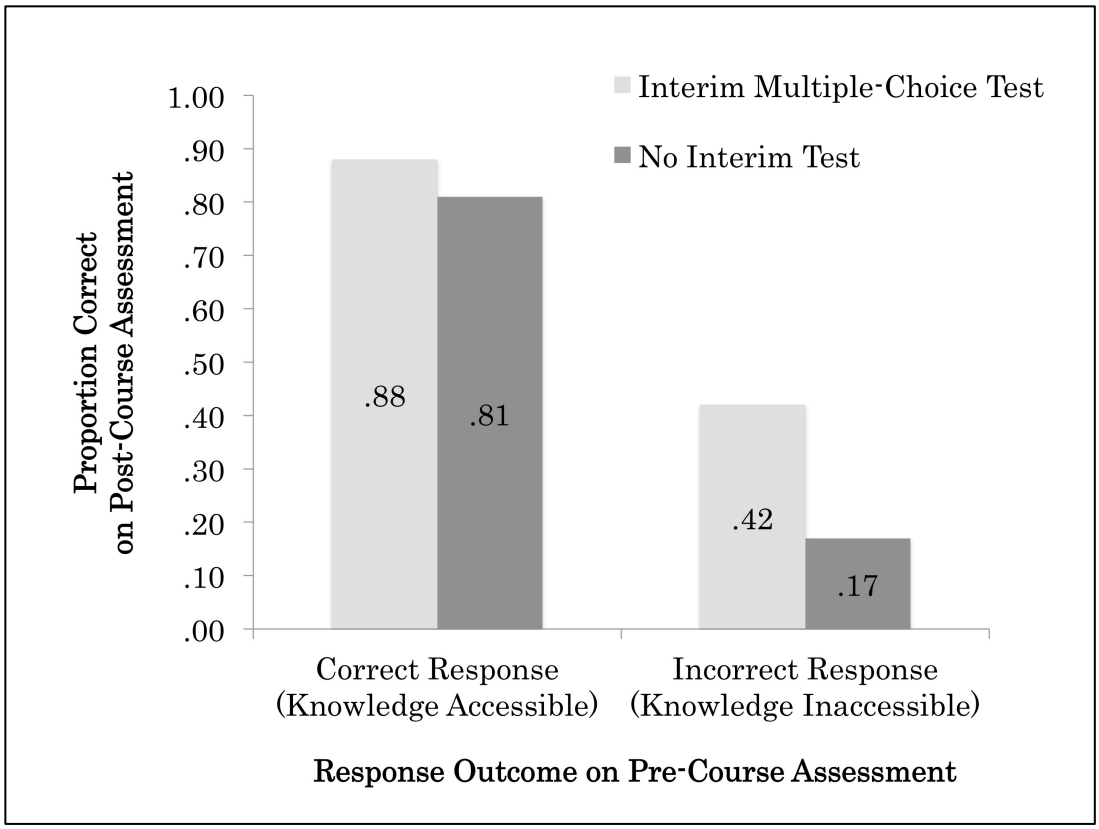

Figure 1. Proportion of correct responses on the post-course assessment as a function of response outcome on the pre-course assessment (correct vs. incorrect) and experimental condition (interim test vs. no test).

References

Bahrick, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General, 120*, 20-33

Berger, S. A., Hall, L. K., & Bahrick, H. P. (1999). Stabilizing access to marginal and submarginal knowledge. *Journal of Experimental Psychology: Applied, 5*, 438-447.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale: Erlbaum.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109*, 204–223.

Brown, R., & McNeil, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325–337.

Cantor, A. D., Eslick, A. N., Marsh, E. J., Bjork, R. A., & Bjork, E. L. (2015). Multiple-choice tests stabilize access to marginal knowledge. *Memory & Cognition. 43,* 193-205.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational research, 66*, 227-268.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381–391.