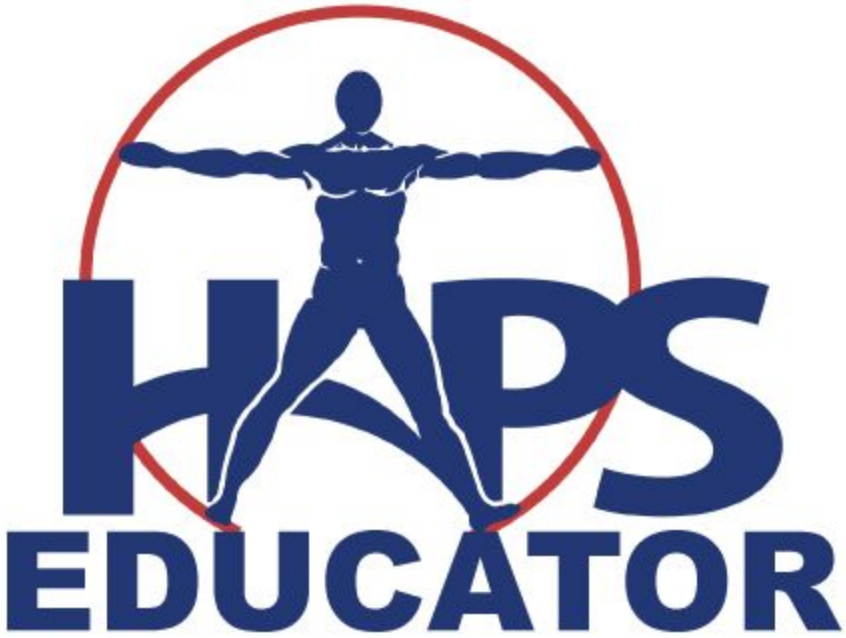
Voluntary Web-Based Self-Assessment Quiz Use is Associated With Improved Exam Performance, Especially for Learners with Low Prior Knowledge.

Jenifer C. Utz* and Matthew L. Bernacki
*Corresponding Author: jenifer.utz@unlv.edu

HAPS Educator. Vol 22 (2), pp. 129-135. Published August 2018.
doi: 10.21692/haps.2018.021



Utz J.C. and Bernacki M.L. (2018). Voluntary Web-Based Self-Assessment Quiz Use is Associated With Improved Exam Performance, Especially for Learners with Low Prior Knowledge. *HAPS Educator* 22 (2): 129-135. doi: 10.21692/haps.2018.021

Voluntary Web-Based Self-Assessment Quiz Use is Associated With Improved Exam Performance, Especially for Learners with Low Prior Knowledge

Jenifer C. Utz, PhD[1] and Matthew L. Bernacki, PhD[2]
[1]School of Life Sciences, College of Sciences, University of Nevada Las Vegas, 4505 S. Maryland Parkway, Las Vegas, NV 89154
[2]Learning Analytics Initiative, College of Education, University of Nevada Las Vegas, 4505 S. Maryland Parkway, Las Vegas, NV 89154
jenifer.utz@unlv.edu, matt.bernacki@unlv.edu

Abstract
This study examined students' voluntary use of digital self-assessment quizzes as a resource for learning in a large anatomy and physiology lecture course. Students (n = 238) could use 16 chapter quizzes and four analogous unit quizzes to rehearse and self-assess knowledge. Most students (75%) engaged in occasional use of self-assessment quiz items; repeated use was uncommon (12%), as was lack of use (13%). Exam performance differed between quiz use groups. Quiz use improved exam performance more among students who entered the course with low prior knowledge of concepts from the prerequisite course. Cumulatively for all students and all exams, repeated self-assessment quiz users significantly outperformed occasional users (+7.5%) and non-users (+11.9%) on course exams. Incorporation of optional learning resources can enhance the learning success of students.
https://doi.org/10.21692/haps.2018.021

Key words: retrieval practice, digital resources, large enrollment, at-risk students, anatomy and physiology

This material is based upon work supported by the National Science Foundation under Grant No. (NSF DRL #1420491). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Introduction
Studies in the United States and abroad indicate that preparatory courses (e.g. Human Physiology, Human Anatomy and Physiology) for students pursuing healthcare professions can be challenging (Higgins-Opitz and Tufts 2013, Hopp 2009, Sturges *et al.* 2016). For example, one study of Human Anatomy and Physiology I students from a small liberal arts college in the southwestern US reported that 43.6% of students (n = 546) earned a W, D, or F at the end of the term (Hopp 2009). An investigation at a southeastern US university documented the grade expectations for over 1200 students enrolled in Human Anatomy and Physiology classes; 65.5% of students overestimated their final course grade (Sturges *et al.* 2016). A study of a first-semester Human Physiology course at an ethnically diverse South African university documented a 54% pass rate for students in a large enrollment (n = 214) course (Higgins-Opitz and Tufts 2013). Improving student learning success in early science courses is a priority for both educators and institutions (Holdren and Lander 2012). Among other major recommendations to the STEM education community, advisory councils recommend that achievement can be increased if educators build useful resources for students and incorporate "active learning" that aligns to learning processes known to improve performance (Holdren and Lander 2012).

Quizzes that can be completed without impact on a course grade are a common learning resource that provides an opportunity for students to engage in "active learning." Active learning is a large category of learning activities that subsume many different processes, including some processes known to promote learning and positively affect performance (Freeman *et al.* 2014, Michael 2006, Prince 2004). One example is students' use of diagrams and figure captions or associated text as a resource to engage in *self-explanation*, wherein the practice of rehearsing one's knowledge orally or in writing improves retention of the knowledge. For example, students tend to perform better on later assessments when they are prompted to self-explain how a biological system works (e.g. the circulatory system, Chi *et al.* 1994). A similar principle can be observed when students engage in *retrieval practice*, a method commonly used to learn factual or declarative knowledge (Karpicke and Blunt 2011). Like self-explanation, students actively engage with targeted content by repeatedly attempting to retrieve answers in response to stimuli.

Students who use retrieval practice or self-explanation methods tend to perform better than those who take more passive approaches like re-reading biology texts (Dunlosky *et al.* 2013). Students who use retrieval practice also tend to

continued on next page

perform better than those who utilize more engaging learning strategies like elaborative concept mapping (Karpicke and Blunt 2011). Both self-explanation and retrieval practice have the added benefit of generating correctness feedback and opportunities to compare students' answers to models. These opportunities enable students to engage in *metacognitive monitoring* (Winne and Hadwin 1998). When students attempt quiz questions on a digital platform, the software can judge their correctness, and thus the quality of their knowledge on topics. Students can monitor this feedback, make judgments about their level of knowledge compared to standards set by course objective, and they can plan their future study accordingly. Active learning strategies, like retrieval practice, are potent and can improve performance, but students tend not to use them in authentic learning environments like their large lecture courses. Research on student learning shows that students tend to opt for minimally effective strategies like re-reading over more effective strategies like retrieval practice and self-explanation (McCabe 2011). This study examined the effect of providing a bank of ungraded digital self-assessment quizzes on student achievement in a large Anatomy and Physiology I lecture course.

The research questions that were investigated include:

1. What percentage of students will make use of the digital quizzes provided to them? How many will use the digital quizzes in a repeated fashion, which might provide benefits associated with retrieval practice and metacognitive monitoring processes?
2. Do students who use digital self-assessment quizzes to differing extents perform differently on course exams?
3. How do the benefits of self-assessment quizzing differ for students with differing levels of prior knowledge?

Methods

Data were collected from students (n = 238) enrolled in a Human Anatomy and Physiology I course at the University of Nevada Las Vegas (UNLV). Two 75-minute lectures were scheduled each week for 15 weeks. Final exams were conducted during the 16th week of the semester. Students attended one 165-minute laboratory session each week. Lecture sections for this course typically contain 180 to 220 students; laboratory sections contain a maximum of 24 students and are typically taught by graduate students. The lecture course has an associated learning management system (LMS; Blackboard Learn) where the instructor provided a variety of resources for students (e.g., chapter learning objectives, lecture presentations, and self-assessment quizzes). Students were introduced to these resources during the first lecture of the semester and utilized them on a voluntary basis thereafter. Course exam structure and content are indicated in Table 1.

All self-assessment quizzes were available to students throughout the entire semester; the Blackboard Learn site opened 1 week prior to the start of instruction and closed 2 weeks after the end of instruction. Quizzes had no time limit and unlimited attempts were allowed. The description stated, "You can use this quiz as often as you wish to test your mastery of the terms and concepts covered in this chapter of the course." Students could use (or not use) the quizzes however they wished. All self-assessment quizzes were composed of a mixture of multiple choice and fill-in-the-blank style questions. The course instructor selected relevant questions from question banks provided by the textbook publisher to build the self-assessment quiz pools. A pool of approximately 35 questions was created for each chapter. Chapter quizzes contained 15 items randomly selected from the pool. Unit quizzes containing 40 items and a 100-item comprehensive quiz covering all chapters were also available. The unit quizzes and the comprehensive quiz pulled questions from the corresponding chapter quiz pools. After submitting a quiz, students received feedback on the correct response as well as the corresponding section of the text. None of the questions utilized for self-assessment quizzes were included on course exams. Course exams were comprised of a mixture of multiple choice questions written by the instructor, test bank questions modified by the instructor, and short answer questions written by the instructor.

After the semester ended, we investigated the effects of self-assessment quiz use and prior knowledge on exam performance. In the context of this study, the phrase prior knowledge is used to describe the knowledge retained from the prerequisite course. Students were grouped by prior knowledge levels into low (n = 84), mid (n = 76), and high (n = 78) terciles based on a 30-item pre-test given at the start of the semester. The pre-test covered general biology content discussed in the prerequisite course. Students in the low prior knowledge group earned pre-test scores in the failing range (< 59%); the mid prior knowledge group earned scores in the C to D range (60 to 79%); the high prior knowledge group earned scores in the A to B range (> 80%). Students were also categorized into groups based on the number of self-assessment quiz items attempted. Students in the "no use" group attempted 0 quiz items over the full semester; students in the "occasional use" group registered between 1 and 400 quiz events; students in the "repeated use" group spanned from 401 through 1478 quiz events. Four hundred events would be observed if a student completed every available self-assessment quiz and registered an answer to each item once.

Before conducting inferential analyses examining differences in exam performance across repeated, occasional, and non-users of self-assessment quizzes, a series of preliminary analyses were conducted to investigate whether the students in these groups differed in their demographics, prior

continued on next page

achievement, or initial motivation in the course. Chi-squared (χ^2) tests and a one-way multivariate analysis of variance (MANOVA) were utilized for the preliminary analyses. After preliminary analyses were completed, student exam scores were submitted to a 3 (Group) x 4 (Time; exams) mixed analysis of variance (ANOVA). The main effect of time is not pertinent here, thus only the main effect of Group is reported to summarily describe the differences in achievement between user groups. Where pertinent, we supplement by reporting contrasts between groups at specific exam time points, and report interactions where groups demonstrate increasing differences in achievement over time. For each comparison, the magnitude of the effect is included. Data were analyzed using statistical software (SPSS); test statistics producing $p < .05$ were considered statistically significant. The institutional review board for social sciences research at the University of Nevada Las Vegas (UNLV) approved this project, IRB #850677, and informed consent was obtained from all participants.

Table 1. Course Exam Structure and Content

	Type of Exam	Exam Structure	Topics Covered
Exam 1	Unit Exam	50 multiple-choice items and 2 short answer items	homeostasis, chemistry, cells, tissues, and integumentary system
Exam 2	Unit Exam	50 multiple-choice items and 2 short answer items	osseous tissue, axial skeleton, appendicular skeleton, and articulations
Exam 3	Unit Exam	50 multiple-choice items and 2 short answer items	muscle tissue, muscular system, and neural tissue
Exam 4	Comprehensive Final Exam	100 multiple-choice items: 60 items on units 1 through 3 and 40 items on unit 4	topics for units 1-3 listed above spinal cord, brain, autonomic nervous system, and special senses

Results

Student Profile

The undergraduate population at UNLV is diverse, and students enrolled in science courses generally reflect the demographics of the larger population (Figure 1). Students in the Anatomy and Physiology course were categorized into groups based on the number of self-assessment quiz events completed (Figure 2).

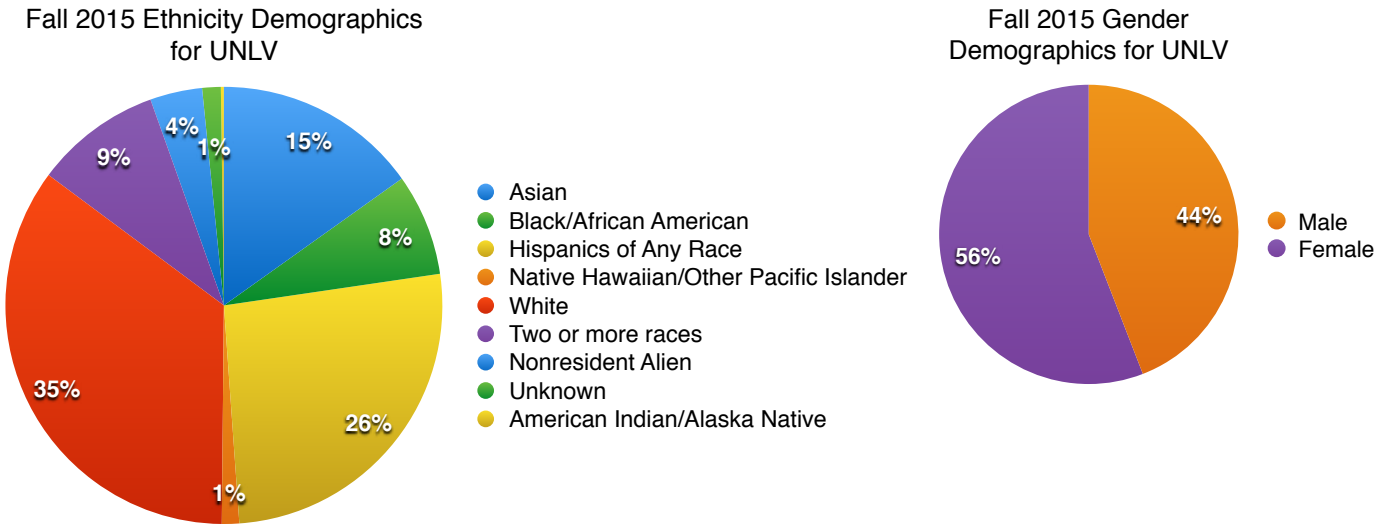


Figure 1. Student Demographics - University of Nevada Las Vegas (UNLV)

continued on next page

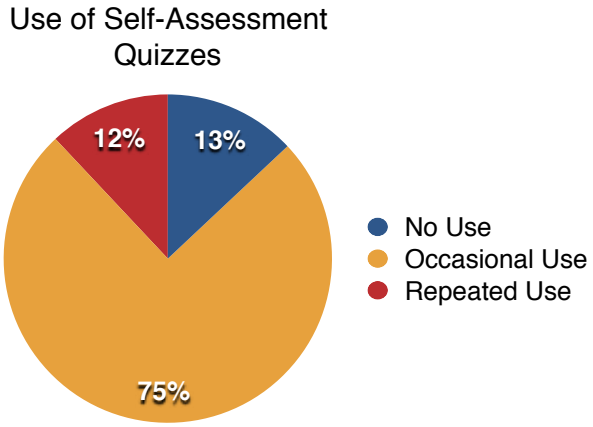


Figure 2. Levels of Voluntary Use of Self-Assessment Quiz Items

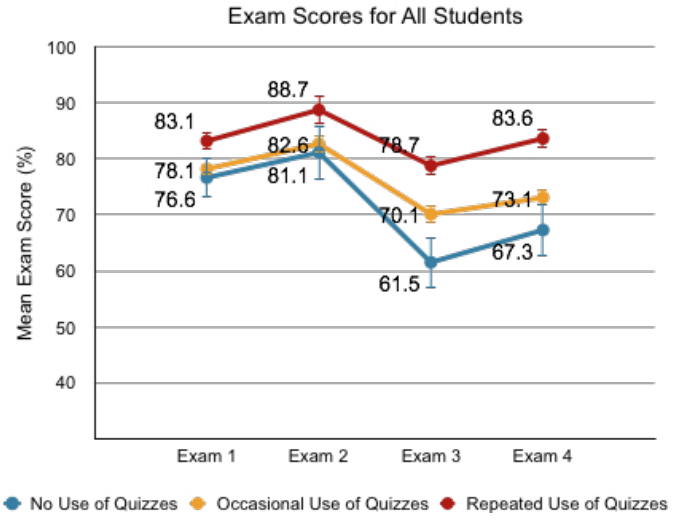


Figure 3. Effect of Self-Assessment Quiz Use on Exam Performance Symbols represent means ± standard error of the mean.

Preliminary Analyses

Preliminary analyses investigated the characteristics of students in the three quiz use groups (none, occasional, repeated). A series of chi-squared (χ^2) tests were conducted to determine whether the three user groups were proportionate in their inclusion of men vs. women, first generation vs. continuing generation college students, or individuals from historically well-represented groups (i.e., Caucasian and Asian-American) vs. underrepresented groups (i.e., Latino/a and African-American). For the whole sample, 12% of students used quizzes repeatedly, 75% percent used them occasionally, and 13% did not use them at all. The χ^2 tests revealed no differences in membership by gender, χ^2 [2, 239] = 0.025, p = 0.998, or generational status, χ^2 [2, 239] = 0.167, p = 0.446. A significant difference in membership by ethnicity indicated that fewer Latino/a and African-American students were repeated quiz users (6%) and more of them were non-users (19%) than would be expected, χ^2 [2, 239] = 6.294, p = 0.043.

A one-way multivariate analysis of variance (MANOVA) was conducted to examine whether self-assessment quiz user groups differed by their prior achievement or achievement motivation for their STEM coursework. At the beginning of the semester, students completed a pretest assessing their retention of knowledge from the prerequisite biology course and completed a battery of motivational scales representing students' self-reported efficacy (Bandura 1977), expectancies and values (Wigfield and Eccles 2000), achievement goals (Elliot and Murayama 2008) for STEM coursework, their perceptions of the values and costs of such coursework, and their general level of academic anxiety and growth mindset orientation. All scales were adopted based on guidelines for their design with specific populations (Bandura 2001) or from prior research with a parallel population (Perez *et al.* 2014) to ensure validity of use. No significant differences were observed across the quiz use groups on this set of measures related to their baseline academic achievement goals and motivation, Wilks' Λ (11, 223) = 1.014, p = 0.445.

Effects of Self-Assessment Quiz Use and Prior Knowledge on Exam Performance

Results of the mixed ANOVA indicated that student use of self-assessment quizzes had a significant effect on course exam performance, F (2, 229) = 5.89, p = 0.003. Post hoc analyses indicated that students who repeatedly used self-assessment quizzes cumulatively scored 7.55% higher on course exams than students who occasionally used quizzes (LSD; p = 0.006) and scored 11.92% higher than students who did not use quizzes (LSD; p = 0.001). Overall, repeated self-assessment quiz users significantly outperformed occasional users and non-users on course exams (Figure 3).

Results of the mixed ANOVA indicated that student prior knowledge level had a significant main effect on course exam performance, F (2, 229) = 18.18, p < 0.001. Overall, students who entered the course with higher levels of prior knowledge outperformed those with lower levels of prior knowledge (Figure 4). Results of the mixed ANOVA indicated an interaction between self-assessment quiz use and prior knowledge level F (4, 229) = 3.51, p = 0.008. Repeated quiz use was associated with greater changes in exam performance for students who entered the course with low levels of prior knowledge (Figure 4A). Students in the low prior knowledge group who repeatedly used quizzes outperformed non-users by nearly 40% on exams 3 and 4, and occasional users outperformed non-users by more than 20% on these exams. Stronger performance on exams 3 and 4 is noteworthy since these exams assess some of the more difficult course concepts (i.e. muscle and neural physiology; Table 1).

continued on next page

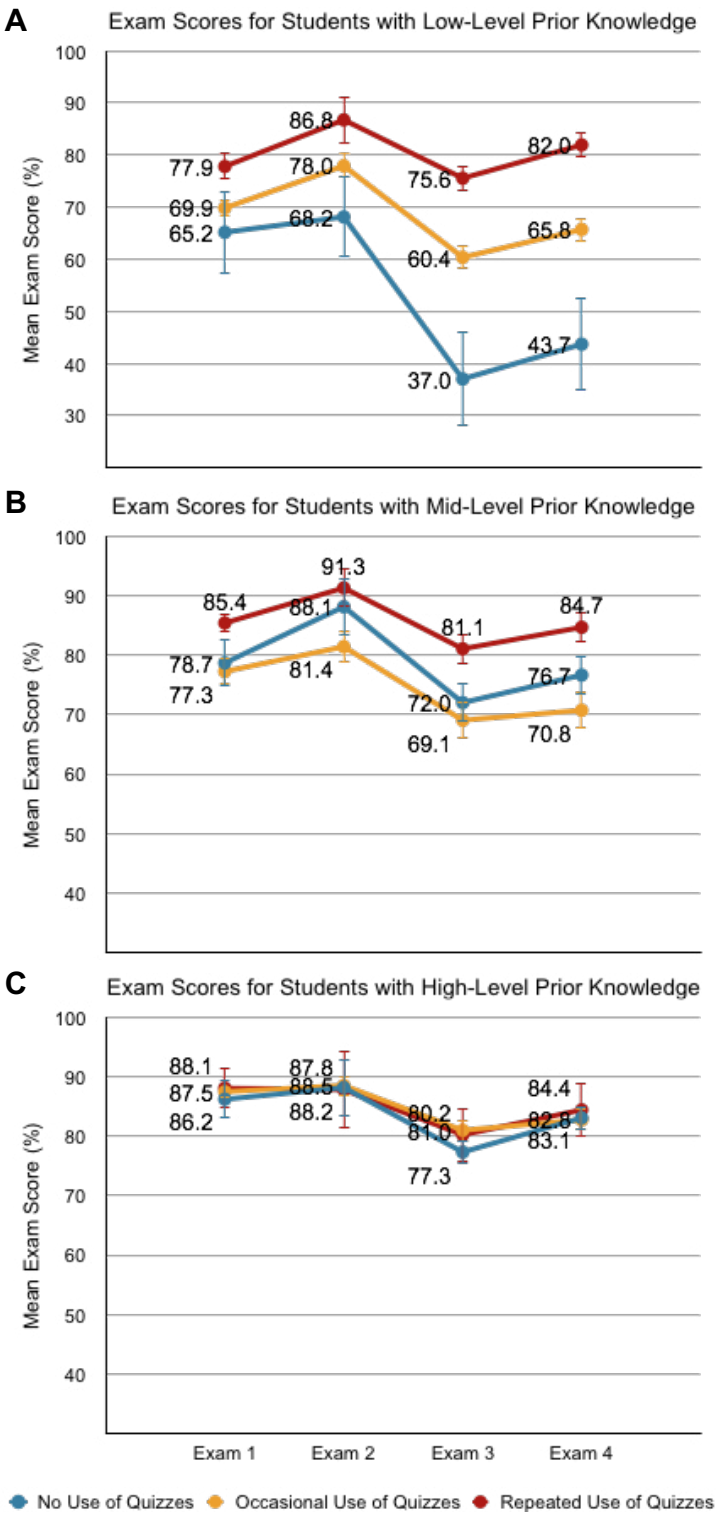


Figure 4. Effects of Prior Knowledge and Self-Assessment Quiz Use on Exam Performance
Symbols represent means ± standard error of the mean.

Discussion

Human Anatomy and Physiology courses are notoriously difficult, and students often enter these courses underprepared. This study investigated the effect of voluntary use of self-assessment quizzes on exam performance. It was documented that digital quiz use substantially improved exam performance, particularly for students with low levels of prior knowledge. This type of course enrichment could benefit students at many diverse institutions across various course sizes and delivery formats (i.e. face to face, hybrid, online). Implementation of this type of self-assessment quiz is aligned with recommendations to improve STEM education through incorporation of evidence-based learning practices (Holdren and Lander 2012). The incorporation of optional quizzes as a learning resource is a reasonable expectation of instructors who can make use of existing resources provided by textbook publishers. Digital quizzes are easily incorporated into LMS course sites; a process that requires limited time investment.

The first research question centered on how students would use *optional* self-assessment quizzes. The effectiveness of retrieval practice via quizzing has been well documented, however many investigations rely on student completion of required quizzes. For example, Norton and Clancy previously documented the effectiveness of digital quizzes in an anatomy and physiology course (2005). In their study, when all students were required to complete two practice quizzes prior to each lab exam, the overall pass rate for the course increased by 8.5% (Norton and Clancy 2005). The Norton and Clancy investigation differs from the current study in a few meaningful ways; all students were required to take multiple practice quizzes, some of the practice quiz questions were included on the subsequent exams, and student subpopulations were not investigated. The present study investigated voluntary use, items on the self-assessment quizzes were not utilized in course exams, and subpopulations of students (i.e. prior knowledge groups) were evaluated. Regarding the level of use, we document that most students (87%) engaged in voluntary use of quizzes that were explicitly intended for self-assessment, with 12% of students repeatedly using the quizzes (Figure 2).

The second research question targeted the relationship between the extent of quiz use and exam performance. The preliminary analyses indicated that students in different quiz use groups did not differ in prior achievement goals or motivation, thus we conclude subsequent differences in exam performance by quiz use groups are likely to have been induced by student learning behaviors (e.g. pattern of quiz use). As might be expected given the opportunities for retrieval practice and metacognitive monitoring, students who repeatedly used self-assessment quizzes earned higher exam scores than other students (Figure 3). However occasional use of quiz items was also beneficial, particularly

continued on next page

on later exams that covered more conceptually challenging material (Figure 3, Table 1). Cumulatively, for all course exams, repeated quiz users averaged 84%, occasional users averaged 76%, and non-users averaged 72%. The preliminary analyses indicated that fewer Latino/a and African-American students were repeated quiz users (6%) and more were non-users (19%) than would be expected. Given these findings, instructors should consider strategies to ensure all students engage with effective learning resources and are aware of the associated educational benefits that correlate with investing time into effective learning practices.

The third research question examined the relationship between prior knowledge level and various levels of quiz utilization. Even with a prerequisite course in place, students begin a new course with vastly divergent levels of knowledge and skill proficiency. Students in the low prior knowledge group earned a pre-test score that would typically be associated with a failing grade (< 59%). Students in this group who did not use self-assessment quizzes averaged 43.7% on the comprehensive final exam whereas their counterparts who repeatedly used the quizzes averaged 82.0% on the final exam (Figure 4A). These gains in exam performance "closed the gap" between these students and their peers who entered the course with high levels of prior knowledge. Students in the high prior knowledge group who repeatedly used the quizzes averaged 84.4% on the final exam (Figure 4C).

Students in the low prior knowledge group who repeatedly utilized self-assessment quizzes experienced the greatest gains in exam performance. In contrast, there was little difference in exam performance for students in the high prior knowledge group. The nature of the quiz items may have contributed to this lack of effect. The quizzes contained questions targeting lower levels (i.e. recall and comprehension) of knowledge whereas the exams contained a mixture of comprehension and application level questions. Levels of knowledge can be described by common schema such as the original and revisions to Bloom's Cognitive Taxonomy (Anderson *et al.* 2001, Bloom *et al.* 1956) or Webb's (1997) Depth of Knowledge, which underlie standards for science learning. A modification that might better support all students would be development of two (or more) levels of self-assessment quizzes. The first level quiz could contain items targeting foundational, declarative knowledge to be recalled (like the quizzes utilized in the present study), and the second level quiz could contain items targeting more advanced levels like applying and analyzing concepts. Such an approach could provide scaffolding for students with low prior knowledge as well as useful learning resources for students with higher levels of prior knowledge.

Increasing the success of vulnerable student populations is a priority for educational and community stakeholders. Incorporation of evidence-based practices, such as low-stakes quizzing to facilitate retrieval practice and metacognitive monitoring, is one path towards improving educational outcomes in STEM disciplines. This investigation documented that most students used optional self-assessment quizzes, repeat users earned higher exam scores, and gains in exam performance were greatest for students who entered the course with a low level of prior knowledge. Students with low prior knowledge may be considered at-risk for failure. Grade point average (GPA) was found to be the strongest predictor of course grades in a study that examined more than a decade of data from 12 life science courses (Creech and Sweeder 2012). Incorporation of optional quizzes is reasonable for instructors and can dramatically benefit students. Students who began with low prior knowledge and repeatedly used self-assessment quizzes ended the semester with a final exam average within 2% of peers who began with high prior knowledge. In sum, instructors are encouraged to consider incorporating optional learning resources that promote retrieval practice and metacognitive monitoring into their courses.

About the Authors

Jenifer Utz is an Associate Professor in Residence in the School of Life Sciences at the University of Nevada Las Vegas (UNLV). Matthew Bernacki is an Assistant Professor in the Learning Sciences and Psychological Studies Program at the University of North Carolina School of Education and consulting director on the UNLV Learning Analytics Initiative. Jenifer teaches a variety of large enrollment introductory biology and human anatomy and physiology courses. Matt teaches courses in learning theory, educational technology, research methods, and statistics.

Jenifer and Matt have collaborated on a variety of STEM education research projects centered on utilization of digital resources and learning management system data to improve students' cognitive and metacognitive skills, their achievement, and their persistence in the STEM majors.

Literature Cited

Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Wittrock MC *et al.* (2001) A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition. White Plains, NY: Longman.

Bandura A (1977) Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.* 84(2): 191-215.

Bandura A (2001) Guide for constructing self-efficacy scales (Monograph). Stanford, CA: Stanford University.

Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR (1956) Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company.

Chi MT, De Leeuw N, Chiu MH, LaVancher C (1994) Eliciting self-explanations improves understanding. *Cogn. Sci.* 18(3): 439-477.

continued on next page

Creech LR, Sweeder RD (2012) Analysis of student performance in large-enrollment life science courses. *CBE – Life Sci. Educ*. 11(4): 386-391.

Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ, Willingham DT (2013) Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychol. Sci. Public Interest* 14(1): 4-58. doi:10.1177/1529100612453266

Elliot AJ, Murayama K (2008) On the measurement of achievement goals: Critique, illustration, and application. *J. of Educ. Psychol.* 100(3): 613-628.

Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014) Active learning increases student performance in science, engineering, and mathematics. *PNAS* 111(23): 8410-8415.

Higgins-Opitz SB, Tufts M (2013) Performance of first-year health science students in a large, diverse, multidisciplinary, first-semester, physiology service module. *Adv. Physiol. Educ*. 38: 161-169. doi:10.1152/advan.00067.2013.

Holdren JP, Lander ES (2012) Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics (Executive Report). Washington, D.C.: President's Council of Advisors on Science and Technology.

Hopp RMP (2009) A success story: chemistry before Anatomy and Physiology I. *HAPS Educator* 14(1): 56-59.

Karpicke JD, Blunt JR (2011) Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*. *331*(6018): 772-775.

McCabe J (2011) Metacognitive awareness of learning strategies in undergraduates. *Mem. Cognition*. 39(3): 462-476.

Michael J (2006) Where's the evidence that active learning works? *Adv. Physiol. Educ*. 30: 159-167. doi:10.1152/advan.00053.2006.

Norton JD, Clancy AN (2005) Use of on-line practice quizzes enhances student performance in anatomy and physiology. *HAPS Educator* 9(2): 10-11.

Perez T, Cromley JG, Kaplan A (2014) The role of identity development, values, and costs in college STEM retention. *J of Educ. Psychol*. 106(1): 315-329.

Prince M (2004) Does active learning work? A review of the research. *J. of Engineering Educ*. 93: 223-231. doi: 10.1002/j.2168-9830.2004.tb00809.x

Sturges D, Maurer TW, Allen D, Gatch DB, Shankar P (2016) Academic performance in human anatomy and physiology classes: a 2-yr study of academic motivation and grade expectation. *Adv. Physiol. Educ*. 40: 26-31. doi:10.1152/advan.00091.2015.

Webb NL (1997) Criteria for alignment of expectations and assessments in mathematics and science education (Research Monograph No. 6). Madison, WI: University of Wisconsin-Madison, National Institute for Science Education.

Wigfield A, Eccles JS (2000) Expectancy-value theory of achievement motivation. *Contemp. Educ. Psychol.* 25(1): 68-81.

Winne P, Hadwin A (1998) Studying as self-regulated learning. In Hacker D, Dunlosky J, and Graesser A (eds.). (pp. 279–306). *Metacognition in Educational Theory and Practice*, Erlbaum: Hillsdale, NJ.

Back to TOC