

Frequent collaborative quiz taking and conceptual learning

Active Learning in Higher Education

2015, Vol. 16(3) 187–196

© The Author(s) 2015

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1469787415589627

alh.sagepub.com

**Ali R Rezaei**

California State University–Long Beach, USA

Abstract

This quasi-experimental study reports on the effectiveness of three assessment strategies for students' performance. The primary goal was to determine whether there are any improvements in students' conceptual learning when a frequent (weekly) quiz is used for grading purposes compared to using midterm and final examinations only. Another goal was to ascertain whether students perform better when they are allowed to work collaboratively on quizzes compared to students taking quizzes individually. The results showed that when the quizzes are open-book, and students have a chance to collaborate (discuss in pairs how to answer the quiz questions), they perform significantly higher in their final examinations and their final projects (an indicator of conceptual learning).

Keywords

assessment strategies, collaborative learning, frequent testing, open-book tests

Assessment strategies in higher education

The recent paradigm shift in higher education from a focus on teaching to a focus on learning has amplified the importance of assessment strategies in many countries (Rust, 2002). According to some educators, learning is so driven by assessment that the form and nature of assessment often swamp the effect of any other aspect of the curriculum (Boud, 1990). Generally, there are two main purposes of student assessment. The first one is called formative assessment which intends to improve the quality of learning, and the second is called summative assessment that concerns the accreditation of knowledge or performance (Boud, 1990). Frequent collaborative testing is introduced in this article as an assessment strategy in which both of these goals are targeted simultaneously.

Frequent assessment could be used as an integral part of teaching to engage students and to enhance their learning (Rezaei and Lovorn, 2010). It is suggested that frequent testing encourages students to stay on task in class, motivates learners to study and review more often, gives teacher a chance to correct students' errors, and gives students the opportunity to figure out what is important and what are they exactly expected to learn (Phelps, 2012). However, we need to learn how to use frequent testing as a way to involve students in meaningful and active learning. Several

Corresponding author:

Ali R Rezaei, 1250 Bellflower Blvd, Long Beach, CA, USA, 90840.

Email: al.rezaei@csulb.edu

approaches have been considered to improve students' learning through frequent testing. Some educators have suggested using open-book tests while others recommend closed-book tests (Gharib et al., 2012). Some have argued that tests are more successful if teachers provide students with their notes, while others believe students should take their own notes (Raver and Maydosz, 2010). An important factor that has not received much attention in the literature is the effectiveness of collaborative test taking. The following literature review synthesizes the literature on frequent testing with an attempt to find the most effective ways in which teachers can engage students in conceptual learning.

Recently, technology has made it possible to design and administer tests much faster and easier than the past. Technology has also facilitated teachers' communication with students and collaboration among students. Consequently, three topics have been given special attention in the literature: first, the effectiveness of frequent testing on students' achievement (Bangert-Drowns et al., 1991; Bařol and Johanson, 2009; Phelps, 2012); second, students' access to teacher's notes or the comparison of open-book, open-note, or open-web with closed-book tests and quizzes (Agarwal and Roediger, 2011; Gharib et al., 2012); finally, the use of PowerPoint and/or class notes in teaching and the comparison of the time (before lecture vs after lecture) of students' access to class notes (Westerkamp et al., 2013). As reflected in the following summary of the literature, the results of these investigations are limited because most of earlier investigations have focused on one strategy at a time.

Frequency of testing

Many studies have been conducted in the past to investigate whether students who are frequently tested learn better than those who are only tested at the middle and/or end of the semester (summative assessment). Summative assessment has been traditionally used in higher education, and still many teachers use them for grading purposes. However, as noted by some educators, summative assessments occur at the end of a semester when there is no time for correction in students' projects or papers after receiving feedback from the instructor (Pokorny and Pickford, 2010). Others have argued that even when timely formative written feedback is provided, it may not be read, may not be understood, and may not be acted upon (Gibbs and Simpson, 2003). Therefore, frequent (daily or weekly) testing and frequent effective feedback has been suggested as a process of ongoing engagement through the provision of opportunities for self-assessment and dialogue (Race, 2008).

With formative assessment through frequent testing, teachers try to urge students to read and practice more. However, if the test or quiz is not graded or does not have marks (points) attached, many students will either not do it at all or only do it in a perfunctory way (Rust, 2002). Teachers also expect that frequent testing and immediate feedback will help both the teacher and students to identify their weaknesses and/or their misconceptions. Through frequent testing and immediate feedback "students generate internal feedback as they monitor their engagement with learning activities" (Nicol and Macfarlane-Dick, 2006: 2). Finally, frequent testing is expected to positively influence students' study habits. However, researchers have not ascertained whether frequent testing actually leads to higher order thinking and deeper understanding (conceptual learning) or simply increases factual knowledge.

Frequent testing is usually not an easy task for either students or instructors. Taking tests can be stressful for many students and designing and grading frequently could also be an overload for teachers. Furthermore, some teachers, particularly in higher education, are not convinced of the positive effects of frequent testing (Bařol and Johanson, 2009). They believe that with frequent testing teachers will not have enough time for instruction and students may focus on how to perform better on the test rather than focusing on learning. Educators have criticized this phenomenon,

well known as “teaching to the test.” For example, Volante (2004) reported that while students’ scores rise when teachers teach closely to a test, learning often does not change. This viewpoint has led some higher education instructors to avoid tests and focus mainly on papers and projects. However, one should note that “teaching to the test” is more likely to happen in high-stake assessments rather than daily or weekly quizzes.

Individual and meta-analysis studies have frequently reported the positive effect of frequent testing on students’ achievement. The positive effects of frequent testing have been observed in a variety of subject matters and grade levels (Bangert-Drowns et al., 1991; Başol and Johanson, 2009; Phelps, 2012). Marsh and Sink (2010) report that long-term retention is generally better after testing as opposed to studying with no testing. Some have compared weekly testing with biweekly testing and others have compared one test with more than one test per semester. Kika et al. (1992) compared two classes taught by the same instructor, the only difference in the classes was the frequency of tests—weekly versus biweekly. They found that mean scores were significantly higher during the weekly testing periods than during the biweekly period. Interestingly, qualitative research studies demonstrate even stronger relationships. For example, Phelps (2012) reported that 93% of qualitative studies have reported on the positive effects of frequent testing while only 1% reported negative effects.

Although several researchers have reported that frequent testing has positive effects on students’ achievement, the results of a meta-analysis by Başol and Johanson (2009) showed that the effectiveness of frequent testing does not differ according to the frequency level used in high-, medium-, and low-frequency group studies. This means that it is not clear how frequent is frequent enough and which subject matter might be appropriate for frequent testing. Başol and Johanson (2009) report that among different subjects, frequent testing was most effective in mathematics. In contrast with the common belief among educators, they report that among the three levels of learning (factual, conceptual, and problem solving), frequent testing was most effective for conceptual learning. Similarly, Hamaker (1986) reported that students’ performance was better on higher order questions than on factual questions. His findings also indicate that when the instructor provides feedback to learners after each test, the effectiveness of frequent testing increases. Regarding the above meta-analyses and individual studies, while it is clear that frequent testing has positive effects in most cases, it is not evident what factors increase or decrease its effectiveness. For example, does it make a difference if the tests are open-book or closed-book?

Open-book testing

The literature on open-book tests is not as rich as the research on test frequency, and the results are not as positive either. Most students prefer open-book or open-note examinations over closed-book examinations (Agarwal and Roediger, 2011). However, as Gharib et al. (2012) report, there is not much agreement among investigators as to which testing style might be more effective in students’ learning and their active engagement in learning. In theory, open-book examinations are designed to encourage students to study more deeply and put less emphasis on memorization (Stalnaker and Stalnaker, 1934). Feller (1994) has argued that open-book examinations are more authentic because people have access to resources they need to perform their job in real life. Researchers have reported different results in comparing open-book and closed-book examinations. Westerkamp et al. (2013) reported that medical students scored significantly higher when preparing for closed-book tests. Similarly, Block (2012) observed that when students were told the examinations were closed-book they came to the test better prepared. However, Gharib et al. (2012) reported that open-book examinations in psychology courses led to better test scores, less test anxiety, and

greater student satisfaction. Similarly, Krarup et al. (1974) reported higher student grades with open-book examinations.

A factor that has not been thoroughly investigated is the long-term effect and long-term retention of material after open-book quizzes. One study showed that while in the short term open-book examinations had positive effects, the effectiveness on long-term retention was not higher than closed-book examinations (Agarwal et al., 2008). Likewise, Patil (2012) did not find a significant difference between the two modes of examinations in terms of student learning over time.

A similar debate exists regarding comparing closed-notes examinations with open-notes examinations (Gharib et al., 2012) and comparing open-web examinations with closed-web examinations (Williams and Wong, 2009). Interestingly, Agarwal and Roediger (2011) concluded that expecting and preparing for open-book examinations may impair long-term retention compared to the situations where students study to prepare for closed-book tests “despite superior initial performance on open-book tests and students’ preference for open-book tests” (p. 836).

Finally, as suggested by Westerkamp et al. (2013), with the increasing rate of online resources available to students, the use of open-book and open-web tests seems inevitable. Nowadays, an important educational task for students is to locate, assess, and properly apply these resources in their projects. According to the authors, students’ ability to locate, assess, and synthesize online/in-print information in a short time period (1–2 hours of examination time) cannot be assessed through closed-book tests.

Collaborative test taking

The merits of cooperative learning for adult learners have been well documented in the literature (Dallmer, 2004). The range of benefits includes engaging students in active learning, promoting self-understanding, encouraging critical thinking, improving interpersonal skills, increasing motivation, and learning about teamwork. Recently, some educators have argued that a vast majority of today’s students will ultimately find work in which they have to interact with other people (Hurren et al., 2006). In order to be successful in the job market in the future, the authors argue that students need to develop an ability to work with others. Yet, not all research has supported collaborative learning activities. For example, Bacon (2005) found that students who completed a project in groups learned less of the project-related content than students who completed a shortened version of the project individually. He argued that some students slack off when they work collaboratively in pairs and that this free riding likely reduces the learning of some students. Similarly, Kennett et al. (1999) concluded that cooperative learning did not facilitate academic performance.

Despite some skepticism about collaborative learning, many instructors are using collaborative projects in their courses because the number of positive reports is significantly more (Billington, 1994). A few studies have indicated the benefits of occasional collaborative test taking (Bloom, 2006; Russo and Warren, 1999). However, systematic and frequent collaborative quiz taking as a teaching strategy rather than as an assessment method has not been investigated. Surprisingly, studies have mainly focused on retention or immediate performance on tests rather than on investigating effects on conceptual learning, critical thinking, and higher order thinking.

In sum, many educators agree on the benefits of frequent testing. While a couple of studies have questioned the effectiveness of frequent testing on conceptual learning, overall it is not yet clear whether frequent testing would simply improve students’ performance on factual tests or whether it would be helpful on higher order thinking and more conceptual questions as well. The literature on the comparison of open-book and closed-book tests is also inconclusive. It is not clear whether open-book examinations are simply offered to reduce students’ stress or whether, in fact, there is

real improvement in students' learning when open-book quizzes are used. More importantly, there is a gap in the literature on frequent collaborative quiz taking as a teaching strategy.

Regarding the aforementioned gap in the literature, it could be concluded that there is a need to examine the effectiveness of frequent testing, open-book quizzes, and collaborative quizzes altogether. There is a need for an experimental or quasi-experimental research to compare "frequent quizzes plus midterm and final examination" with the traditional "midterm and final examination only" assessment strategy. There is also a need to see whether students' collaboration on a test has a short-term or a sustainable effect. It is proposed that if students are given a chance to discuss and collaborate in pairs as they answer the quiz questions, they would perform better both on weekly quizzes and on their final examinations as well as their final projects (an indicator of deep understanding and higher order thinking). Regarding the benefits of collaborative learning as discussed in the literature review, it is hypothesized that if students are allowed to discuss and work together, they will perform better at the end of the semester in comparison with an individual quiz-taking situation.

Methodology

A total of 288 students participated in this study. Various groups of participants took a "Quantitative Research Methods" course in 12 different sections during 5 years from fall 2009 to fall 2014. All participants were from California State University, Long Beach, United States. The age range of the participants was from 24 to 52 years. Some students took the course in regular semesters (Fall, Spring) and other groups took the course in Summer session. Regular semesters take 16 weeks and students meet only once a week. In summer sessions, however, students meet twice a week and the semester is only 6 weeks long.

The same instructor (the author) taught all sections of the course using the same textbook and the same table of contents in the same room. All classes were held in a computer lab where students could take the quizzes on a computer and had access to all their notes as well as the instructor's PowerPoint slides. A midterm, a final examination, and a research project were required for all sections of the course. The midterms and finals had a similar format and the same level of difficulty, but they covered different chapters of the textbook. For the final project, students were expected to choose a research topic and write a 10-page research proposal.

This study compared three phases of teaching using a different assessment strategy in each phase. Each phase consisted of three to five sections of the course being taught by the instructor at that time period. It might be argued that these sections were not comparable in terms of students' characteristics and demography. However, since multiple sections of this course were studied in each phase, it was unlikely that the students in one phase were inherently and significantly different from the other two phases.

In the first four sections of the course (phase 1), the researcher used traditional assessment: students were assessed through a midterm and a final examination and a final project but no quizzes were given. Students in these sections received lectures about each chapter of the book and thereafter participated in an activity related to that chapter. The questions on the examinations were a combination of factual and conceptual questions. For factual questions, students were asked to show their understanding of research terminology, and for conceptual questions, the instructor provided a scenario in which students were asked to propose the most appropriate approach to answer the research question(s).

In the next three sections of the course (phase 2), the instructor offered a quiz after each lecture. The quizzes were weighted only 30% of the total points given for the course. To ensure the equivalency of the quizzes in multiple sections of the course, the same test bank was used in all sections.

Table 1. Descriptive statistics on students' final performance (average of final examination and final projects).

	N	Mean	Standard deviation	Standard error
No Quiz	102	83.1176	6.81189	0.67448
Frequent Quiz	62	86.9194	5.04829	0.64113
Collaborative Quiz	124	90.2097	4.28511	0.38481
Total	288	86.9896	6.27893	0.36999

Table 2. ANOVA comparison of students' final performance under the three assessment strategies.

	Sum of squares	df	Mean square	F	Significance
Between groups	2815.235	2	1407.618	47.198	0.000
Within groups	8499.733	285	29.824		
Total	11314.969	287			

ANOVA: analysis of variance.

For each quiz, a random of 20 questions were selected from the test bank. These quizzes were open-book and open-notes, but students were not allowed to work collaboratively on the test. Everything else in this phase was similar to the first phase of the study, including the final projects, midterms, and final examinations.

In the last five sections (phase 3) of the study, students took a quiz right after the lecture on a chapter. These quizzes were open-book and open-notes, and students were allowed to either work individually on the test or work in pairs to discuss and answer the questions together. Almost all students chose to do the test in pairs. They were not allowed to work with the same partner on different quizzes. Everything else was similar to the second phase of the study including the final projects, midterms, and final examinations. It should be noted that the midterm and final examinations as well as final projects had to be done individually, not collaboratively, in all three phases of this study.

Results

One-way analysis of variance (ANOVA) was used to compare students' learning in the three phases of teaching. The average of students' scores on the final examination and their final project, labeled as "final performance," was used to make this comparison. The midterms and student scores on weekly quizzes were not used in this comparison. Table 1 shows the descriptive results, including the means and standard deviations of students' scores on final performance. Table 2 shows the results of the analysis of variance, and Table 3 shows the results of the post hoc test (Scheffe's test). As shown in these tables, students' "final performance" increased significantly from phase 1 (No Quiz) to phase 2 (Frequent Quiz) as well as from phase 2 to phase 3 (Collaborative Quiz).

During the second and third phases of the study, students took weekly quizzes. In order to show students' progress during the semester, Table 4 compares students' average scores in the "first two quizzes" with their average scores on the "last two quizzes," in phase 2 and phase 3.

This table shows that in both cases, students' performance on weekly quizzes increased. Table 5 shows that in both phase 2 and phase 3, this increase was statistically significant.

Table 3. Post hoc test result to compare each assessment strategy.

(I) Teaching type	(J) Teaching type	Mean Difference (I – J)	Standard error	Sig.
No Quiz	Frequent Quiz	–3.80171	0.87944	0.000
	Collaborative	–7.09203	0.73000	0.000
Frequent Quiz	No Quiz	3.80171	0.87944	0.000
	Collaborative	–3.29032	0.84943	0.001
Collaborative Quiz	No Quiz	7.09203	0.73000	0.000
	Frequent Quiz	3.29032	0.84943	0.001

Table 4. Comparing students' scores at the beginning and the end of the semester.

	N	Minimum	Maximum	Mean	Standard deviation
Phase 2					
First quizzes	62	77.00	100.00	87.4839	5.66769
Last quizzes	62	79.00	96.00	90.1290	3.71741
Phase 3					
First quizzes	124	73.00	93.00	84.0726	4.35456
Last quizzes	124	84.00	100.00	92.4758	3.15059

Table 5. Test of significant difference between the first and the last two quizzes.

	Mean	Standard deviation	Standard error	t	df	Significance (two-tailed)
Phase 1	–2.64516	3.23960	0.41143	–6.429	61	0.000
Phase 2	–8.40323	3.04417	0.27337	–30.739	123	0.000

Discussion and conclusion

It was proposed that if students take quizzes more frequently, they will perform better in comparison with students who simply take a midterm and a final examination. Comparison of students' performance in the first with the second phase showed that when students were tested frequently, their performance improved significantly. Regarding the majority of studies in the literature, this result was expected to a large extent. However, what was interesting was that frequent testing not only improved students' performance in the short term (as reflected in their progress through quizzes) but also improved their deeper and more sustainable understanding, as reflected in their final examination (that addressed mostly higher order thinking and scenario-based problem solving) and in their final project (developing a quantitative research proposal). Therefore, frequent testing was not only helpful in factual learning but also in higher order and critical thinking.

Furthermore, the results showed that frequent testing reduced the individual differences initially observed among students. This means that, as shown in Table 4, both the range and the standard deviation of scores decreased in the final stages of the course. This finding—frequent testing as an equalizer—has not been reported in previous studies on frequent testing.

While no systematic observation was initially planned as the methodology of this study, informal observation of students in phases 2 and 3 confirms earlier studies suggesting that open-book weekly quizzes lead to better test scores, less test anxiety, and greater student satisfaction. This finding was also evident in students' evaluations of the instructor at the end of the last few semesters. The results of this study disagree with some earlier studies (e.g. Agarwal et al., 2008) questioning the effectiveness of open-book examinations on deep understanding and higher order thinking. Considering students' high performance on both their final examination and their final project (which was essentially a higher order and critical thinking activity), this study shows a sustainable, meaningful, learning after frequent open-book testing.

It was also predicted that if students were allowed to work together collaboratively in a test, this would increase their performance not only on that particular test but also on the final projects and examinations taken individually. The results strongly supported this hypothesis. Students in the third group performed significantly higher on their final performance in comparison with the first and the second group.

Regarding the fact that the final performance was done individually and it was composed of a research proposal and a conceptual (higher order thinking) final examination, it is concluded that collaborative learning leads to sustained and meaningful learning. Some have reported that although collaborative work produces higher examination scores, this happens at a cost of increased confidence for groups' wrong answers (Puncochar and Fox, 2004). Students' confidence was not directly measured in this study. However, the instructor tried to rectify students' misconceptions by discussing the questions that were consistently answered incorrectly by students, after they completed the test.

In the first phase of this research, it was noticed that students had a hard time understanding the lectures, particularly when multiple new concepts were included in the lecture. Therefore, in the next phase, students were asked to read the text before the lectures and were told that they will be given a quiz right after the lecture. Reading the chapter before the lecture served to scaffold the new concepts. It is interesting to note that initially this method was not fully successful. Many students came to class unprepared for the first few weeks. However, after a few weeks, they realized that if they did not read the chapter, they would not understand the lecture and/or they would not do well on the quizzes. In the third phase, gradually, students learned how to read the chapters, highlight the confusing parts, and ask the right questions during the lecture. The number of students' questions in class was not measured systematically. However, it was observed that both the number and the quality of students' question increased in comparison with the first phase where no weekly test was given and the second phase where the test was taken individually.

Some educators have argued that in collaborative activities some students might slack off, or not participate or contribute to the discussions (Bacon, 2005). This is particularly possible if students have a specific partner for the whole semester. In this case, it is possible that some students take advantage of their partners and enjoy the free ride. In order to address such concerns, in the third phase, the instructor had students switch their partners for each test. This policy allowed students to have discussions with a different student each time and to have the opportunity to hear different perspectives and different problem-solving strategies.

This study was not true experimental research. The main limitation of the study was lack of random sampling and random assignment. It could be argued that students taking the course at different times are not comparable. Therefore, any differences among the three groups could be attributed (at least to some extent) to individual differences. The other limitation of this study is that it was limited to only one course. It might be argued that different courses and different contexts might lead to different conclusions.

While this study clearly shows the benefits of frequent testing and collaborative quiz taking, there are still many educators skeptical of these assessment strategies (Başol and Johanson, 2009). We need further research in other disciplines to examine the effectiveness of these assessment strategies. Particularly, more controlled experimental research is needed to examine the effectiveness of open-book and collaborative quiz taking on students' stress and their studying habits.

As reflected in the literature, many students and teachers believe that frequent testing is stressful for students (Russo and Warren, 1999). This study, however, shows that if the test is open-book, if the teacher urges students to read the book before the lectures, and if students get a chance to collaborate (discuss) how to answer the quiz, it leads to positive, meaningful, and sustainable effects. It should be noted that in phase 3 of this study, students were allowed to answer the quiz questions in pairs. This policy proved to be successful both quantitatively and qualitatively. Observing students having meaningful and productive discussions about every single quiz question was very reassuring. Students did not appear to be stressed out before, during, or after the quiz. This type of teaching (assessment) was also rewarding for the instructor, as reflected in students' anonymous evaluations of the instructor. If this policy works for a research method course, a course that is not usually the students' most favorite one, it should most certainly work for courses that students like much more.

If you have tried frequent testing but stopped it due to students' stress or their complaints, if you are skeptical about the effectiveness of open-book tests, or if you think students may not study for an open-book or a collaborative test, this research may help you to rethink about some of your assessment strategies.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Agarwal P and Roediger HL (2011) Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory* 19(8): 836–52.
- Agarwal P, Karpicke J, Kang S, et al. (2008) Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology* 22(7): 861–76.
- Bacon DR (2005) The effect of group projects on content-related learning. *Journal of Management Education* 29(2): 248–67.
- Bangert-Drowns RL, Kulik JA and Kulik CC (1991) Effects of frequent classroom testing. *The Journal of Educational Research* 85(2): 89–99.
- Başol G and Johanson G (2009) Effectiveness of frequent testing over achievement: A meta-analysis study. *International Journal of Human Sciences* 6(2): 99–121.
- Billington R (1994) Effects of collaborative test taking on retention in eight third-grade mathematics classes. *The Elementary School Journal* 95(1): 23–32.
- Block R (2012) A discussion of the effect of open-book and closed-book exams on student achievement in an introductory statistics course. *PRIMUS* 22(3): 228–38.
- Bloom D (2006) Collaborative testing: A second chance to learn. *Academic Exchange Quarterly* 10(3): 265–9.
- Boud D (1990) Assessment and the promotion of academic values. *Studies in Higher Education* 15(1): 185–200.
- Dallmer D (2004) Collaborative test taking with adult learners. *Adult Learning* 15(3–4): 4–7.
- Feller M (1994) Open-book testing and education for the future. *Studies in Educational Evaluation* 20(2): 235–8.

- Gharib A, Phillips W and Mathew N (2012) Cheat sheet or open-book? A comparison of the effects of exam types on performance, retention, and anxiety. *Psychology Research* 2(8): 469–78.
- Gibbs G and Simpson C (2003) Does your assessment support your students' learning? *Learning and Teaching in Higher Education* 1(1): 3–31.
- Hamaker C (1986) The effects of adjunct questions on prose learning. *Review of Educational Research* 56(2): 212–42.
- Hurren BL, Rutledge M and Gavrin AB (2006) Team testing for individual success. *Phi Delta Kappan* 87(6): 443–7.
- Kennett DJ, Young AM and Berrill DP (1999) Is cooperative learning effective for high achieving entrance students? Implications for policy and teaching resources. *Journal of Research and Development in Education* 33(1): 27–35.
- Kika FM, McLaughlin TF and Dixon J (1992) Effects of frequent testing of secondary algebra students. *Journal of Educational Research* 85(3): 159–62.
- Krarup N, Naeraa N and Olsen C (1974) Open-book tests in a university course. *Higher Education* 3(2): 157–64.
- Marsh EJ and Sink HE (2010) Access to handouts of presentation slides during lecture: Consequences for learning. *Applied Cognitive Psychology* 24(5): 691–706.
- Nicol DJ and Macfarlane-Dick D (2006) Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education* 31(2): 199–218.
- Patil BM (2012) Comparative study of open book exam and close book exam for VIIIth Std. *Golden Research Thoughts* 2(6): 1–3.
- Phelps RP (2012) The effect of testing on student achievement, 1910–2010. *International Journal of Testing* 12(1): 21–43.
- Pokorny H and Pickford P (2010) Complexity, cues and relationships: Student perceptions of feedback. *Active Learning in Higher Education* 11(1): 21–30.
- Puncochar JM and Fox PW (2004) Confidence in individual and group decision making: When “two heads” are worse than one. *Journal of Educational Psychology* 96(3): 582–91.
- Race P (2008) *Make Learning Happen: A Guide for Post-Compulsory Education*. London: SAGE.
- Raver SA and Maydosz AS (2010) Impact of the provision and timing of instructor-provided notes on university students' learning. *Active Learning in Higher Education* 11(3): 189–200.
- Rezaei AR and Lovorn M (2010) Reliability and validity of rubrics for assessment through writing. *Assessing Writing* 15(1): 18–39.
- Russo A and Warren SH (1999) Collaborative test taking. *College Teaching* 47(1): 18–21.
- Rust C (2002) The impact of assessment on student learning: How can the research literature practically help to inform the development of departmental assessment strategies and learner-centered assessment practices? *Active Learning in Higher Education* 3(2): 145–58.
- Stalnaker JM and Stalnaker RC (1934) Open-book examinations. *The Journal of Higher Education* 5(3): 117–20.
- Volante L (2004) Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy* 35. Available at: <http://www.umanitoba.ca/publications/cjeap/articles/volante.html>
- Westerkamp AC, Heijne-Penninga M, Kuks, JBM, Cohen-Schotanus J (2013). Open-book tests: Search behavior, time used and test scores. *Medical teacher* 4: 330-332.
- Williams JB and Wong A (2009) The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book open-Web exams. *British Journal of Educational Technology* 40(2): 227–36.

Author biography

Ali R Rezaei is a professor at California State University, Long Beach. During the last 20 years, he has been teaching and doing research at City University of New York and the University of Calgary, Canada. Computer applications in education, research methods, statistics, and assessment (evaluation) are among the courses that he teaches. His research interests are assessment, the Internet, and social networks. Address: California State University–Long Beach, CA, USA. [email: al.rezaei@csulb.edu]