

ON STUDENTS' PERCEPTION OF A MULTI-SCHEME ASSESSMENT METHOD

Ambrose Leung¹ and Cheryl A. Kier²

Abstract

The discipline of economics has the reputation of being abstract and difficult to understand. Students' perceptions about the course and its assessment methods may affect their learning strategies and outcomes. A flexible grading scheme (in terms of weights assigned to different assessment components) may help to reduce anxiety about the course, leading to improved performance and higher satisfaction. The present study reports on the implementation of such a flexible grading scheme. Students in five introductory economics classes received their final course grade based on one of three grading schemes that gave them the highest mark. An 11-item satisfaction survey assessed students' perceptions about the flexible marking scheme. Students expressed a preference for flexibility over a more traditional, fixed marking scheme. This has implications for other courses that create anxiety among students and has the potential to improve students' experiences at post-secondary institutions.

Key Words: assessment; undergraduate teaching; students' learning

JEL Classification: A22, I23

Introduction

The study of economics provides students with a set of useful and relevant tools that enables them to make better decisions in every facet of their lives. The main purpose of a course in principles of microeconomics is to teach students how to weigh benefits against costs to achieve the best possible solution to any problem. The discipline of economics, however, has the reputation of a "dismal science" (Perksy, 1990, p.165); an area of study replete with abstract concepts and often perceived as difficult to understand. Borg and Shapiro (1996), for example, comment that students more often hate rather than love economics. Recently, economists have started to question whether adjustments in teaching styles to better accommodate students with different learning styles can improve student achievement and attitudes in introductory economics courses (Boatman, Courtney, & Lee, 2008; Leung, McGregor, Sabiston, & Vriliotis, 2014). Other studies have explored the effects of factors such as class size (Becker & Power, 2001; Kennedy & Siegfried, 1997), instructor effectiveness (Shmanske, 1988; Watts & Bosshardt, 1991), and teaching methods (Emerson & Taylor, 2004; Frank, 1997) on student performance. Most of this literature found mixed evidence on these modifications, revealing a need for more research.

¹ Ambrose Leung is the author for correspondence, associate professor of economics, Department of economics, justice, and policy studies, Mount Royal University, 4825 Mount Royal Gate SW, Calgary, Alberta, Canada, T3E 6K6, e-mail: acleung@mtroyal.ca, phone: (403)-440-8515, fax: (403)-440-6815.

² Cheryl Kier is an associate professor of psychology, Centre for social sciences, Athabasca University, 1 University Drive, Athabasca, Alberta, Canada, T9S 3A3, email: cherylk@athabascau.ca, phone: (403)-289-0408.

Studies from the education literature suggest that students' perceptions about assessment methods could affect their learning strategies and outcomes (Drew, 2001). In this literature, the word "assessment" traditionally describes the processes applied to evaluate students' understanding of the instructional materials. Over the years, the meaning of assessment has broadened to include a wide range of educational activities in the teaching and learning process, referred to as assessment for learning (AfL). The concept of AfL has been an important force in the discussion of education policies.

There does not appear to be a common definition of AfL. AfL can be described as assessment activities and strategies that promote and improve students' learning (William, 2011). William pointed out frequent classroom testing as one useful AfL strategy, because frequent testing has a strong relationship with better student performance. Frequent testing also helps reduce students' stress and boost students' liking of the course (Bangert-Drowns, Kulik, & Kulik, 1991; Dempsey & Perkins, 1993).

Research on testing finds that tests often promote the retention of learned information (Butler & Roediger, 2007; Rickards, 1979; Runquist, 1986) and that the content and format of tests affect how students approach studying and learning for a course (Biggs, 1996; Ha & Kapoor, 2015). Students' learning is adversely affected by assessment methods that are perceived to be inappropriate (Struyven, Dochy, & Janssens, 2005), but learning improves when students perceive assessment as a means to make them accountable for their own learning (Brown & Hirschfeld, 2008).

A literature review by Bangert-Drowns, *et al.* (1991) showed that performance improved when students were given a larger number of shorter tests as opposed to fewer, longer tests. Furthermore, students held more favorable opinions of their classes when they were more frequently tested. This implies that frequent tests may help to foster an environment more conducive to learning and a mutually respectful classroom environment. Frequent testing may discourage students from cramming for exams and reduce test anxiety (Butler & Roediger, 2007; Dempster & Perkins, 1993).

Other studies compared the advantages and drawbacks of different assessment methods. This literature differentiates between surface learning and deep learning. Surface learning refers to learning with limited personal engagement and understanding of the material (Struyven *et al.*, 2005). Surface learners often use routine memorization to complete learning tasks. Deep learning, in contrast, refers to thoroughly understanding the subject matter, leading to much preferred high quality learning outcomes.

Although post-secondary students tend to prefer exams in multiple-choice rather than essay format, essay questions foster the preferred deep learning, while multiple-choice exams encourage surface learning (Birenbaum, 2007). The literature on teaching economics, however, has shown that carefully written multiple-choice questions can test for in-depth understanding (Buckles & Siegfried, 2006). Furthermore, multiple-choice and essay questions measure different dimensions of knowledge, so it is optimal to ask both types of questions on economics exams (Becker & Johnston, 1999; Chan & Kennedy, 2002).

The literature on assessment considers both multiple-choice and essay exams inferior to less conventional assessment methods such as self-assessment, peer assessment, computer-based assessment, and portfolios (Alquraan, 2012). Traditional assessment methods have been criticized for not offering students sufficient flexibility and control over the assessment process (Irwin & Hepplestone, 2012).

While students appear to favor new and innovative assessment methods that improve learning outcomes (Bevitt, 2014; Sternberg, Penn, & Hawkins, 2011), these methods tend to be more costly to implement than traditional paper-and-pencil exams (Alquraan, 2012). Furthermore, many modern methods, such as portfolios and oral examinations, are often impossible for larger class sizes. This implies that finding other ways to improve flexibility in course assessment would be valuable.

Aside from the high costs and difficulties in applying alternative assessment methods, the traditional assessment processes is often associated with anxiety. Two types of students suffer test anxiety (Birenbaum, 2007). One type knows and understands the course material, but suffers interfering thoughts that restrict the process of retrieving relevant information during tests. The other type lacks effective study strategies to acquire the necessary knowledge to tackle a test. Regardless of the type, a more flexible grading scheme - in terms of weights assigned to different assessment components - may reduce anxious feelings about the course and lead to improved performance and higher student satisfaction.

This study adopted a flexible grading scheme based on frequent testing and assignments in order to realize some of the benefits of alternative assessment methods (i.e., reduce student stress and increase their enjoyment) while maintaining student learning and retention in an unpopular principles of microeconomics course.

A Flexible Multi-Scheme Assessment Method

The present study introduced a multi-scheme assessment method in an attempt to improve student motivation and attitudes by using an AfL strategy that encourages deep learning and limits anxiety while remaining feasible for classrooms with limited resources. The principles of microeconomics course was chosen to test this method, because it is notorious as course dreaded by students and has a low average course mark. For example, at the institution where data were collected, the percentage of students who received grades D, F, and W (withdrawn from course) was reported as roughly 15% overall, whereas the DFW rate for Principles of Microeconomics is consistently among the “top 20” at well above 20% (personal communication, November 22, 2016).

Furthermore, the course is required or strongly recommended for many first-term university students as one of their first university courses. The low course average is partly due to the typical grading scheme that consists of a relatively small weight on homework assignments (10 to 20%), with the remaining course grade determined by performance in two midterm exams (20 to 30% each) and a final exam (30 to 40%). Therefore, poor performance on any one of the exams has a significant negative impact on overall course performance. With heavy weighting on in-class testing, many students express stress and anxiety about learning the course material.

To compound the problem of heavy weighting on tests, few students consistently perform at a high level throughout the course. Some students perform well in the beginning and taper off toward the end as the course materials become more complex. Other students require more time to get used to the course materials and perform better at the end compared to the beginning. In some cases, students suddenly “click” with the material only at the very end after performing poorly throughout the term. Given that principles of microeconomics is known to have a low course average, low levels of student satisfaction, and high levels of student anxiety, it was an appropriate course in which to test a flexible multi-scheme assessment method with more frequent testing.

Using the guidelines in Gibbs and Simpson (2004), this course used two strategies for assessment in higher education that support students' learning: 1) a sufficient number of assessed tasks to capture adequate study time and effort as well as to engage students in productive learning activities; and 2) sufficient timely feedback relevant to the criteria for success in the course. Flexibility in assigning weights to different assessment components is achieved by giving more tests.

Frequent testing provides students with more practice as well as more feedback on their understanding of the course material (Gholami, 2014). Frequent testing is a way to promote student learning by encouraging students to engage continuously in the course material. Learning can also occur during testing (Chan, 2016; Richland, Kao, & Kornell, 2008). It can help prevent procrastination and last minute studying right before the final exam (Anthis & Adams, 2012; Kerdijk, Cohen-Schotanus, Mulder, Muntinghe & Tio, 2015). Evidence shows that students enjoy frequent testing and believe they learn more from the course as a result (Roediger & Karpicke, 2006). Frequent testing also allows assigning smaller weights to each test, reducing student anxiety associated with less frequent and bigger exams. This in turn helps to boost test performance (Ali Zarei, 2015). Schrank (2016) found that frequent testing enhanced students' experience and satisfaction with an introductory sociology course that the author described as popular and "widely admired" (p. 118).

As with other formative tasks, frequent testing provides students with feedback, so they can assess how well they understand the material and determine topics for review (Lavery, Bauer, Kortemeyer, & Westfall, 2012). It can also improve students' motivation (Domenech, Blazquez, de la Poza, & Muñoz-Miquel, 2015). Furthermore, feedback helps students to develop deep learning (Holmes & Papageorgiou, 2009).

This study examines whether the adoption of a multi-scheme assessment method, incorporating frequent testing and feedback, could change the attitude and experience of students towards the principles of microeconomics course. Our hypotheses are:

The multi-scheme assessment method will

- (1) reduce students' anxiety associated with the course
- (2) enhance students' satisfaction from the course
- (3) affect students' study strategies, and
- (4) provide better incentive for students to learn the course material

This study assessed students' attitudes toward the course in which the final course grade was determined by one of three grading schemes that gave the student the highest grade. The grading scheme replaced the two midterm exams with five quizzes that were shorter in length and covered less material. One-fifth of the course grade was allocated to computer homework assignments that provided students with instant detailed feedback. These five sets of homework assignments served as practice for each of the five quizzes. The assignments were intended to make students accountable for their own learning before each quiz (Brown & Hirschfeld, 2008). All assignments and tests contained both multiple-choice and essay questions. These included computational and graphical analysis to ensure testing of different dimensions of knowledge and to increase opportunities for students to show their understanding of the materials through different test formats.

Three assessment schemes were offered with each student's final course grade determined by the scheme that offered the highest grade. Students did not need to decide on

which scheme they would use, as their marks were derived automatically from a pre-programmed formula. The three assessment schemes are summarized below:

	Scheme 1	Scheme 2	Scheme 3*
Computer Lab Work	20%	20%	20%
Quizzes	4 (out of 5) @ 15% each =60%	2 (out of 5) @ 15% each =30%	4 (out of 5) @ 20% each =80%
Final exam	20%	50%	0%
Total	100%	100%	100%

*Scheme 3 only applies to students who attended all quizzes on the scheduled days and time.

Scheme 1 applied to students who showed consistent effort and performance over the course. The weight over each assessment component in Scheme 1 is relatively even (15% to 20% each). Scheme 2 is tailored to students who prefer less frequent in-class testing, (e.g., due to a busy work schedule), and to those who tend to show continuous improvement over time. Scheme 3 is designed for students who perform better with shorter tests that cover a smaller quantity of material on each test. To ensure that all students were tested on all course materials, scheme 3 was only an option to students who wrote all five quizzes, because this scheme allowed students to opt out of the final exam. Furthermore, since the computer homework assignments covered all course materials, all schemes ensured that students were responsible for learning all the course material.

Of the 165 students who received a grade for the course, 4.2% (7 students) ended up with Scheme 1 (20% final exam), 8.5% (14) ended up with Scheme 2 (50% final exam), and 87.3% (144) ended up with Scheme 3 (0% final exam).

From the students' perspective, the flexibility of the assessment scheme offered each of them a better chance to obtain a higher course grade relative to fixed marking schemes by accommodating different study preferences/habits and personal schedules. From a pedagogical perspective, the hope is that the more flexible assessment scheme reduces student anxiety and aversion towards the course, and provides students with added motivation to devote more effort toward studying. Improvement in student attitude and experience towards the course may also enhance student learning and retention of the course material.

Participants

One-hundred forty-seven students completed satisfaction surveys about the flexible course scheme, once at the beginning of the course and once at the end. Two different instructors used the scheme, resulting in two classes taught by one instructor and three classes taught by the second instructor. Each class had approximately 35 students. We did not gather information about sex or age from participants, but the University Records Office reported that of 537 students registered in principles of microeconomics courses in the Fall of 2015, most were first year students (N=461; 85.8%) and male (N=299; 55.7%) (personal communication, May 13, 2016).

Measures

A student satisfaction survey was administered at two time points. The survey consisted of 11 questions, eight of which contained a 5-point Likert scale, along with three open-ended

questions. An additional question asked students which of the three schemes they thought would result in their highest mark. The survey questions included, “How much difference do you think having a variety of marking schemes will make to your satisfaction with the course?” and “How much do you think your studying will be affected by having a variety of marking schemes?” The surveys took five to ten minutes to complete.

Procedure

Ethical approval was provided by the Research Ethics Boards of both authors’ institutions. To avoid any conflict of interest and to prevent students from feeling an obligation to complete the surveys, the surveys were distributed in the classrooms by people who were not associated with the university. The course instructors were absent during the survey so that students did not feel coerced to participate.

Results

Chi-square analyses were conducted for each of the eight Likert scale items to compare scores at the beginning versus the end of the semester. See Table 1 for these results.

TABLE 1: Chi-square statistics³ of the eight Likert scale survey items

Question number	Frequency counts for answering 1,2, or 3 beginning of semester; N=146 (%)	Frequency counts for answering 1,2, or 3 end of semester N=147 (%)	Chi-square results for differences between beginning and end of semester
1	5 (3%)	5 (3%)	n. a.
2	18 (12%)	5 (3%)	Chi-Square (2) = 4.72, p<.10
3	36 (23%)	35 (24%)	n. a.
4	135 (92%)	138 (94%)	n. a.
5	108 (74%)	114 (78%)	Chi-square (4) = 18.8, p<.001.
6	18 (12%)	9 (6%)	Chi-square (2) = 4.96 p<.10
7	36 (25%)	21 (14%)	Chi-square = 29.53 (4), p<.001.
8	103 (71%)	106 (72%)	n. a.

For each question we compared scores at one end of the continuum (1, 2, and 3) to scores at the other end of the continuum (4 and 5). However, because we reversed the polarity for some of the questions to avoid response sets (e.g., International Encyclopedia of the Social Sciences, n.d.), the definitions of the anchor points are unique for each question. These definitions are covered below as we describe the results for each question in more detail. Several of the questions reached the ceiling (92% or higher agreement) at which no differences were found.

³ The formula used to compute the Chi-square statistic is $\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, where O and E represents the observed and expected values respectively.

Taken as a whole, the results of both the quantitative data and the open-ended answers of the satisfaction survey indicate that students were satisfied with the flexible marking scheme. We present evidence for each question of the survey.

1: *How much do you like the idea of having a variety of marking schemes?*

Don't like Very Much					Like Very Much
1	2	3	4		5

At both the beginning and the end of the semester, 97% of students replied that they highly or very highly liked the idea of having a variety of marking schemes.

2: *How much difference do you think having a variety of marking schemes will make to your satisfaction with the course?*

Not Very Much Difference					Very Much Difference
1	2	3	4		5

The Chi-square analysis shows that 88% of students indicated they thought the flexible marking scheme would affect their satisfaction in the course at the beginning of the semester while 97% believed this at the end of the semester. There is a marginally significant trend for a difference between beginning and end of term when scores 1-3 are collapsed and compared against scores 4 and score 5. The trend is toward very much difference in their satisfaction in the course, as many more students chose 5, but fewer students chose 4 or 1 at the end of the semester.

3: *How much difference do you think the different marking schemes will make to your performance on the course?*

Not Very Much Difference					Very Much Difference
1	2	3	4		5

At the beginning of the semester, 23% of students reported they believed the flexible marking scheme would not make much difference to their course performance. At the end of the semester, 24% reported that they did not think the flexible marking scheme made much difference to their performance. The difference is not significant.

4: *How confusing do you think it is to have a variety of marking schemes?*

Not Very Confusing					Very Confusing
1	2	3	4		5

There was no significant difference between the beginning and the end of the semester in terms of confusion over the flexible marking scheme. At the beginning, 92% of students believed it was not confusing while 94% believed this at the end of the semester.

5: *How highly would you recommend this variety of marking schemes for other courses?*

Highly Recommend					Not Recommend at All
1	2	3	4		5

There was a small, but statistically significant difference in recommendations of the flexible marking scheme from beginning to end of semester based on the Chi-square analysis. At the beginning 74% were likely to recommend a flexible marking scheme vs. 78% at the end of the semester, Chi-square (4) = 18.8, $p < .001$. Students moved from selecting high scores (4 & 5) at the beginning of the semester toward selecting lower scores at the end of the semester, indicating an increased likelihood of recommending flexible marking schemes for other courses.

6. *If you were able to choose, how likely would you be to choose a section that had a variety of marking schemes over a section that only offered one marking scheme?*

Not Very Likely Very Likely
 1 2 3 4 5

At the beginning of the semester, 88% of students indicated they were likely to choose a section of the course that had a flexible marking scheme compared to 94% at the end of the semester. This is marginally significant for a change in likelihood when answers 1-3 are collapsed, Chi-square (2) = 4.96, $p < .10$. Students moved their answers more from 1 to 5. This indicates a shift toward greater likelihood to choose a course with a variety of marking schemes.

7. *How much do you expect to enjoy this course as compared to courses that do not use a variety of marking schemes?*

Enjoy Much Less Enjoy Much More
 1 2 3 4 5

At the beginning of the semester, 75% of students believed they would enjoy the course more since it had a flexible marking scheme. This compares to 86% who believed this at the end of the semester. This was a significant difference, Chi-square (4) = 29.53, $p < .001$. The shift from beginning of term to end of term was from lower numbers to higher numbers. This indicates that students believed the flexible marking scheme helped them to enjoy the course more than they expected to.

8. *How much do you think your studying will be affected by having a variety of marking schemes?*

Very Much Change Not Very Much Change
 1 2 3 4 5

There was no perceptible change in the number of students who believed their studying would be affected by having a flexible marking scheme instead of a fixed one. At the beginning of the semester, 71% believed this compared to 72% at the end of the semester.

8a. *In what ways has your studying has been affected?*

This open-ended question revealed that 55 of 147 students (37%) said they studied harder/more/more consistently. Six students said they studied less or put in less effort, because they did not need to worry about marks when there were so many opportunities to improve.

9. What do you like best about having different marking schemes?

At the end of the semester, 14 students (out of 147; 10%) mentioned less stress or pressure was a benefit of the flexible marking scheme. Sample answers included, “the flexibility it allows, and the reduction in stress caused by being able to measure marks in different ways,” “have to study more for the quizzes and show up to every class so I do not have extra exam stress”, and “it makes the course less stressful, and easier to get the higher mark”.

Not having to take a final exam was a motivating factor for a number of students; 42 out of 147 students (29%) mentioned this. Typical comments included, “I felt like I had to be much more prepared for each quiz to be able to use a marking scheme with a lower value for the final” and “I have kept up with the course work because I am motivated not to write the final as I am not good at writing finals. This helped me understand the course material better.”

10. What do you think is the worst feature about having different marking schemes?

Twenty three of 146 students (15.8%) at the beginning of the semester and 19 of 147 students (12.9%) at the end of the semester said that having different options of marking schemes could be confusing. Sample comments include, “people may be confused,” “confusing at times and tends to cause me to overthink,” and “somewhat confusing but once you’re used to it everything is fine.”

11. Which marking scheme do you think will lead to your highest mark?

<i>Scheme 1</i>	<i>Scheme 2</i>	<i>Scheme 3</i>
<i>(20% final)</i>	<i>(50% final)</i>	<i>(no final)</i>

In the beginning of the semester, 109 of 146 or 74.7% of students thought Scheme 3 with no final exam would lead to the highest mark, compared to 118 of 147 (80.3%) students at the end of the semester. Very few students thought Scheme 2, with a 50% weight on the final exam, would lead to the highest mark both in the beginning (5.5% of students) and at the end (4.1% of students) of the semester.

Discussion

The purpose of this paper was to explore whether an assessment scheme increased students’ motivation and reduced their anxiety in principles of economics courses. We also wanted to know whether the multi-scheme assessment method would affect students’ study strategies. The evaluation scheme provided flexibility in terms of weights assigned to various assessment components in the course. Overall, students appear to like the flexibility offered by the multi-scheme assessment method very much. The results further show that this attitude was consistent at the beginning and at the end of the semester.

As for level of anxiety, the written comments provided some support for students’ enjoyment of the multi-scheme assessment method. They frequently mentioned reduced stress or anxiety (17.8% and 27.2% in the beginning and end of semester, respectively) as benefits of the flexible scheme.

Studying strategies appear to change only moderately based on the Likert scale instrument. About two-thirds of students (67% and 63% in the beginning and end of semester, respectively) gave some indication in their written comments that the multi-scheme assessment method motivated them to keep up with the course throughout the term.

While the students certainly seemed to enjoy the flexibility offered by the multi-scheme assessment method, the current study was only tentatively able to identify the reasons for this. In written comments, students claimed they appreciated the opportunity to avoid the final exam (35.6% and 39.5% in the beginning and end of semester, respectively); believed the flexible scheme increased performance and success in the course (30.1% and 31.3%); and provided flexibility (36.3% and 40.8%). However, we are unable to determine which components of the scheme account for their satisfaction.

For example, the possibility of avoiding the final exam is a component of the assessment method. Such a feature can increase motivation to study consistently throughout the semester and it can reduce stress associated with taking the course. Further research should investigate whether merely removing a final exam is sufficient to obtain such high levels of satisfaction. Although there is debate in the literature about the value of final exams (e.g., Glass, Ingate, & Sinha, 2013; Khanna, Badura Brack, & Finken, 2013; Szpunar, Mcdermott, & Roediger III, 2007), research on frequent testing and retention suggests that students may learn and remember as much from courses using this method as from those with a heavily weighted final exam.

Nevertheless, the impact of the flexible assessment method on student learning remains unclear. For example, will a student be less likely to remember the course material without writing the final exam, or by writing a final exam that carries only a small weight in the course grade? It is possible that students feel less incentive to study (and thus remember the material) if they expect the exam will not count for very much. The data used here, unfortunately, does not constitute a large enough sample to analyze the issue of information retention.

Of the 165 students who received a grade, only 25 wrote the final exam, as the vast majority managed to complete all the assignments and quizzes. Four, seven, and fourteen students ended up using the final exam to account for 0% (average mark on the final exam = 72.82), 20% (average mark = 44.53), and 50% (average mark = 61.5) of the course grade, respectively. Within this small sample, students who wrote, but did not end up using, the final exam to account for anything in their course grade earned an average mark that was much higher than that of their counterparts. If the final exam indicates how well students retain course information, then this may imply that students who keep up with studying throughout the term remember the material best.

However, the seven students who ended up with the final exam counting for 20% had a lower average mark than the 14 students who used the final exam for 50% of their course grade. This may imply that students have less incentive to study for the final exam when the weight of the final exam is low. Student learning and information retention in relation to the flexible assessment scheme is an important issue that requires further research.

Conclusion

This study sheds light on the potential of multi-scheme assessment methods to reduce the long-standing negative perception associated with principles of economics courses. Given that students indicated a strong preference for a flexible marking scheme compared to a fixed one, we recommend that courses in other disciplines investigate its use. Although previous research suggests that frequent tests improve student performance (Bangert-Drowns et al., 1991; Domenech, et al., 2015; Lavery et al., 2012), future research could examine whether students' course marks are affected by these changes in motivation and anxiety.

References

- Ali Zarei, A. 2015. "On the effectiveness of quizzes on L2 idioms learning." *Iranian Journal of Language Testing*, 5(1): 60-77.
- Alquraan, M.F. 2012. "Methods of assessing students' learning in higher education: An analysis of Jordanian college and grading system." *Education, Business and Society: Contemporary Middle Eastern Issues*, 5(2): 124-133.
- Anthis, K. & L. Adams. 2012. "Scaffolding: Relationships among online quiz parameters and classroom exam scores." *Teaching of Psychology*, 39(4): 284-287.
- Bangert-Drowns, R.L., J.A. Kulik, & C-L C. Kulik. 1991. "Effects of frequent classroom testing." *Journal of Educational Research*, 85(2): 89-99.
- Becker, W.E. & C. Johnston. 1999. "The relationship between multiple choice and essay response questions in assessing economics understanding." *The Economic Record*, 75(231): 348-357.
- Becker, W.E. & J.R. Powers. 2001. "Student performance, attrition, and class size given missing student data." *Economics of Education Review*, 20(4): 377-88.
- Bevitt, S. 2015. "Assessment innovation and student experience: A new assessment challenge and call for a multi-perspective approach to assessment research." *Assessment & Evaluation in Higher Education*, 40(1): 103-119.
- Biggs, J. 1996. "Enhancing teaching through constructive alignment." *Higher Education*, 32(3): 347-364.
- Birenbaum, M. 2007. "Assessment and instruction preferences and their relationship with test anxiety and learning strategies." *Higher Education*, 53(6): 749-768.
- Boatman, K., R. Courtney, & W. Lee. 2008. "'See how they learn': The impact of faculty and student learning styles on student performance in introductory economics." *The American Economist*, 52(1): 39-48.
- Borg, M.O., & S.L. Shapiro. 1996. "Personality type and student performance in principles of economics." *Journal of Economic Education*, 27(1): 3-25.
- Brown, G.T.L. & G.H.F. Hirschfeld. 2008. "Students' conceptions of assessment: Links to outcomes." *Assessment in Education: Principles, Policy & Practice*, 15(1): 3-17.
- Buckles, S. & J.J. Siegfried. 2006. "Using multiple-choice questions to evaluate in-depth learning of economics." *Journal of Economic Education*, 37(1): 48-57.
- Butler, A.C. & H.L. Roediger, III 2007. "Testing improves long-term retention in a simulated classroom setting." *European Journal of Cognitive Psychology*, 19(4/5): 514-527.
- Chan, N. & P.E. Kennedy. 2002. "Are multiple-choice exams easier for economics students? A comparison of multiple-choice and "equivalent" constructed-response exam questions." *Southern Economic Journal*, 68(4): 957-971.
- Chan, S.H. 2016. "Testing effect in a college class." *HKICSS/ACGEBF/GCEP Conference Proceedings*, Paper presented at the Higher Education Forum, Hong Kong, June 8-10, 2016, Hong Kong (pp. 266-271), Hong Kong: HEF.
- Dempster, F.N. & P.G. Perkins. 1993. "Revitalizing classroom assessment: Using tests to promote learning." *Journal of Instructional Psychology*, 20(3): 197-203.
- Domenech, J., D. Blazquez, E. de la Poza, & A. Muñoz-Miquel. 2015. "Exploring the impact of cumulative testing on academic performance of undergraduate students in Spain." *Educational Assessment Evaluation and Accountability*, 27(2):153-169.
- Drew, S. 2001. "Student perceptions of what helps them learn and develop in higher education." *Teaching in Higher Education*, 6(3): 309-331.

- Emerson, T.L.N. & B.A. Taylor. 2004. "Comparing student achievement across experimental and lecture-oriented sections of a principles of microeconomics course." *Southern Economic Journal*, 70(3): 672-693.
- Frank, B. 1997. "The impact of classroom experiments on the learning of economics: An empirical investigation." *Economic Inquiry*, 35(4): 763-769.
- Gholami, V. & M.M. Moghaddam. 2013. "Effect of weekly quizzes on students' final achievement score." *International Journal of Modern Education and Computer Science*, 5(1), 36-41.
- Gibbs, G. & C. Simpson. 2004. "Conditions under which assessment supports students' learning." *Learning and Teaching in Higher Education, Issue 1, 2004-05*, p. 3-31.
- Glass, A.L., M. Ingate, & N. Sinh. 2013. "The effect of a final exam on long-term retention." *The Journal of General Psychology*, 140(3): 224-241.
- Ha, T. & V. Kapoor. 2015. How assessment, motivation and beliefs about learning shape students' learning approaches. Paper presented at International Conference: Assessment for Learning in Higher Education 2015, University of Hong Kong, Hong Kong, May 15, 2015.
- Holmes, K., & G. Papageorgiou. 2009. "Good, bad and insufficient: Students' expectations, perceptions and uses of feedback." *Journal of Hospitality, Leisure, Sport and Tourism Education*, 8(1): 85 - 96.
- Irwin, B. & S. Hepplestone. 2012. "Examining increased flexibility in assessment formats." *Assessment & Evaluation in Higher Education*, 37(7): 773-785.
- Kennedy, P.E., & J.J. Siegfried. 1997. "Class size and achievement in introductory economics: Evidence from the TUCE III data." *Economics of Education Review*, 16(4): 385-394.
- Kerdijk, W., J. Cohen-Schotanus, B.F. Mulder, F.L.H. Muntinghe, & R.A.Tio. 2015. "Cumulative versus end-of-course assessment: effects on self-study time and test performance." *Medical Education*, 49(7): 709-716.
- Khanna, M.M., A.S. Badura Brack, & L.L. Finken. 2013. "Short- and long-term effects of cumulative finals on student learning." *Teaching of Psychology*, 40(3): 175-182.
- Laverty, J.T., W. Bauer, G. Kortemeyer, & G. Westfall. 2012. "Want to reduce guessing and cheating while making students happier? Give more exams!" *The Physics Teacher*, 50(9): 540-543.
- Leung, A., M.A. McGregor, D. Sabiston, & S. Vriolotis. 2014. "VARK learning styles and student performance in principles of micro- vs. macro-economics." *Journal of Economics and Economic Education Research*, 15(3): 113-120.
- Persky, J. 1990. "Retrospective: A dismal romantic." *Journal of Economic Perspective*, 4(4): 165-172.
- "Response Sets." *International Encyclopedia of the Social Sciences*. Retrieved December 20, 2016 from Encyclopedia.com: <http://www.encyclopedia.com/social-sciences/applied-and-social-sciences-magazines/response-sets>
- Richland, L.E., L.S. Kao, & N. Kornel. 2008. "Can unsuccessful tests enhance learning?" In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2338-2343). Austin, TX: Cognitive Science Society.
- Rickards, J. 1979. "Adjunct postquestions in text: A critical review of methods and processes." *Review of Educational Research*, 49(2): 181-196.

- Roediger, H.L. III & J.D. Karpicke. 2006. "The power of testing memory: Basic research and implications for educational practice." *Perspectives on Psychological Science*, 1(3): 181-210.
- Runquist, W.N. 1986. "The effect of testing on forgetting of related and unrelated associates." *Canadian Journal of Educational Research*, 40(1): 65-76.
- Schrank, Z. 2016. "An assessment of student perceptions and responses to frequent low-stakes testing in introductory sociology classes." *Teaching Sociology*, 44(2): 118-127.
- Shmanske, S. 1988. "On the measurement of teacher effectiveness." *Journal of Economic Education*, 19(4): 307-314.
- Szpunar, K.K., K.B. Mcdermott, & H.L. Roediger III. 2007. "Expectation of a final cumulative test enhances long-term retention." *Memory & Cognition*, 35(5): 1007-1013.
- Siegfried, J.J., & F. Rendigs. 1979. "Research on teaching college economics: A survey." *Journal of Economic Literature*, 17(3): 923-69.
- Sternberg, R.J., J. Penn, & C. Hawkins. 2011. *Assessing College Student Learning: Evaluating Alternative Models, Using Multiple Methods*. Washington, DC: Association of American Colleges and Universities.
- Struyven, K., F. Dochy & S. Janssens. 2005. "Students' perceptions about evaluation and assessment in higher education: A review." *Assessment & Evaluation in Higher Education*, 30(4): 331-347.
- Watts, M., & W. Bosshardt. 1991. "How instructors make a difference: Panel data estimates from principles of economics courses." *Review of Economics and Statistics*, 73(2): 336-40.
- Wiliam, D. 2011. "What is assessment for learning?" *Studies in Educational Evaluation*, 37(1): 3-14.