

Online Test Presentation Modes, Student Self-perceived Learning Styles, and Student Performance on Factual/Conceptual and Applied Problems

Yuejin Xu, Murray State University

Abstract

Online courses have created a different testing environment. A test in an online setting can be easily presented in multiple modes (i.e., all at once, one at a time). However, there are no established guidelines to date on how the test presentation modes should be used in an online test. Using a sample of sixty-five undergraduate students, this study examined if test presentation modes (i.e., all at once, one at a time), student self-perceived learning styles (i.e., surface style, deep style), and types of test problems (i.e., factual/conceptual questions, applied questions) have an effect on student achievement in four unit tests. Findings from one-way MANOVA and repeated measures ANOVAs revealed that test presentation modes alone did not influence student test performance. However, when the variables of student self-perceived learning styles and types of test problems were added, students with a surface style scored significantly higher on the factual/conceptual problems in the one-at-a-time test presentation mode than in the all-at-once test presentation mode. No significant differences were found for students with a deep style in test performance based upon test presentation modes and types of questions. This study suggests test presentation modes can be set according to students' learning styles. Students with a surface style may benefit from the one-at-a-time test presentation mode over time. Furthermore, it is recommended that technical issues and test security should be considered in determining an optimal test presentation mode.

Keywords: student achievement, online test, test presentation mode, learning style

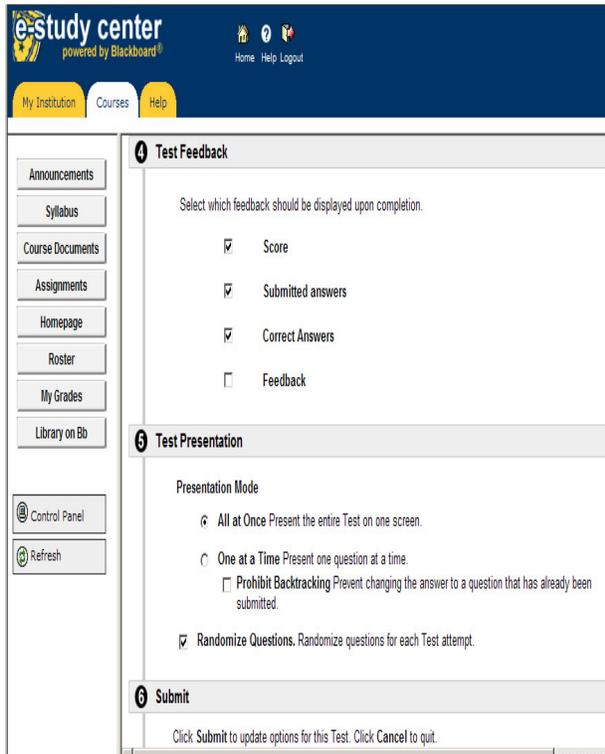
Introduction

More and more courses in public schools and universities are using online learning technology, such as Blackboard® course management system. Online courses (no or minimal face-to-face class meetings) and web-enhanced courses (regular face-to-face class meetings) have created a different testing environment. A test in an online setting (such as in Blackboard®) can be easily presented in three modes (Figure 1), namely, all at once (present the entire test on one screen), one at a time (present one question at a time), and one at a time with no backtracking (present one question at a time, preventing changing the answer to a question that has already been submitted). The instructor manual of Blackboard® briefly describes the function of each option; however, it did not offer any suggestions on

when each option would be optimally used. A search of literature further reveals that there are no established guidelines on choosing the presentation mode in an online test for online course instructors and online test designers.

In addition, a majority of research on online tests or computer-based tests focused on how test items can be selected and presented from large test databases to make a reliable and valid test (Barrada, Olea, Ponsoda, & Abad, 2008; Doong, 2009; Finkelman, Nering, & Roussos 2009; Meijer & Nering, 1999; Pastor, Dodd, & Chang, 2002; Revuelta & Ponsoda, 1998; Stocking & Lewis, 1998; van der Linden & Veldkamp, 2007; Veldkamp & van der Linden, 2008). Only a few studies have examined the impact of presentation modes

Figure 1. Test Presentation Modes in Blackboard®



on student test performance (Truell, Zhao, & Alexander, 2005).

Truell, Zhao, and Alexander (2005) examined business students' scores based on "the settable test item exposure control interface format" (p. 31). The phrase of "settable test item exposure control interface format" is synonymous with "test presentation mode" used in the present paper. Their results indicated there was "no significant difference in postsecondary business student scores or test completion times based on settable test item exposure control interface format" (p. 31). However, they also recommended that "it would be prudent to conduct additional research in a variety of settings" (p. 38). Given that Truell, Zhao, and Alexander's sample was from the business setting, it is worthwhile to replicate this study using a sample from a different field, and to consider additional

potential factors affecting student test performance, such as student learning styles, and types of test problems.

Student Learning Styles

Learning styles are approaches to or ways of studying and learning (Woolfolk, 2008). Of the many models of learning styles (Coffield, et al., 2004), Snow, Corno, and Jackson's model (Snow, Corno, & Jackson, 1996) was adopted in the present study for its simplicity and ease of assessment. Snow, Corno, and Jackson (1996) identified two types of learning styles: surface-processing approach vs. deep-processing approach. Students who take a surface-processing approach (surface style) focus on ways to acquire and maintain information through audio/visual aids, note taking, repetition, and memorization. They also tend to depend on authority as the source of learning. Students who have a deep-processing approach (deep style) focus more on understanding, making sense and meaning of the material through self-reflection, discussion, and practice (such as hands-on activity). Deep style students are more likely to be self-motivated. They are likely to depend on the self as the source of learning. Learning styles are related to student achievement (Coffield, et al., 2004; Dunn, 1987; Dunning, 2008; Rogers & McNeil, 2009; Stahl, 2002; Terregrossa, Englander, & Wang, 2010).

Types of Test Problems

Multiple-choice format is a common type of objective testing. It has been widely used to measure student achievement. Based on Bloom's cognitive taxonomy (Bloom et al., 1956), most multiple-choice questions can be classified into three types, namely, factual, conceptual, and applied. In this study, factual and conceptual types of questions are combined because both reflect the lower level in the taxonomy.

Factual/conceptual questions are used to measure one's knowledge or understanding of given material without relating it to another situation. In contrast, applied questions demand higher level of thinking from test-takers. Applied questions reportedly help students to improve their learning (Winne, 1979) and lead to more accuracy in measuring student performance.

The purpose of this study was to examine if test presentation modes (i.e., all at once, one at a time), student self-perceived learning styles (i.e., surface style, deep style), and types of test problems (i.e., factual/conceptual questions, applied questions) have an effect on student achievement. Specifically, the study addressed three research questions:

1. Does test presentation mode influence student test performance over time?
2. Does test presentation mode influence the surface style students' test performance on factual and applied problems over time?
3. Does test presentation mode influence the deep style students' test performance on factual and applied problems over time?

Method

Participants and Setting

The participants in this study were 65 undergraduate students at a public university in the South. They were recruited from two sections of one undergraduate core course (Psychology of Human Development) in fall 2008 semester. Most of them were female, white, and in their twenties. One section of students was randomly assigned to the all-at-once test presentation condition (a total of 35 students). The other section was assigned to the one-at-a-time test presentation condition (a total of 30 students). Even though students enrolled in those sections by their own preferences and availability of spaces in those sections, no marked differences were

found among students in the two sections in terms of age, gender, and ethnicity. The two sections received the same amount and type of instruction (web-enhanced) from the same course instructor. The two sections shared the same course syllabus.

Measures

Student test performance was assessed by four unit tests. Each test consisted of 40 multiple-choice questions adapted from a large test bank accompanying the course textbook. Each test was worth 40 points. All tests were given through the Blackboard® course management system in a proctored lab environment during the regular class meeting times. All students were able to complete each test using less time than the given 75 minutes.

The unit tests contained factual/conceptual and applied problems. Except for Test 1, which contained 28 factual/conceptual types of problems, and 12 applied type of problems, each of the other three tests (Test 2, Test 3, and Test 4) contained 20 factual and 20 applied problems.

Student self-perceived learning styles were measured by students' written responses to a short essay question, "How would you describe your way of learning?" Two raters independently coded each student's written response into either the deep style or surface style. Cohen's Kappa statistics was performed to determine agreement between the two raters. The inter-rater reliability was found to be $Kappa = .84$ ($p < .001$), indicating a very good level of agreement. Differences between raters were further resolved through peer discussion.

Procedures

Students were required to submit a short essay assignment to Blackboard® during the first week of class. One of the

short essay questions was “How would you describe your way of learning?” Students’ written responses were collected for analysis to determine their self-perceived learning style. One student did not submit her written response to this question. The student was excluded from analysis when the self-perceived learning style variable was added.

Students took all the tests in a computer lab. All tests were given online through the Blackboard® in a close-book, close-notes proctored environment. Test 1 was presented in one-at-a-time-without-backtracking mode for all students in the two sections. Test 2, 3 and 4 were presented differently between the two sections. One section of students was randomly assigned into the condition where test items were presented all at once. The other section was assigned into the condition where test items were presented one at a time. Institutional Review Board (IRB) approval was obtained. Willing students signed consent forms to allow the investigator to analyze their written responses and test scores.

Data Analysis

1. Does test presentation mode influence student test performance over time?

A one-way multivariate analysis of variance (MANOVA) was conducted to examine whether there were any pre-existing differences between the two sections in 1) overall test 1 scores, 2) factual/conceptual problems scores in test 1, and 3) applied problems scores in test 1.

Three repeated measures ANOVAs were conducted using time as the within-subjects variable, comparing student overall test scores in test 2, test 3, and test 4, factual problems scores in the three tests, applied problems scores in the three tests, respectively. Test presentation mode (all at once, one at a time) was the between-

subjects variable in the three repeated measures ANOVAs.

2. Does test presentation mode influence the surface style students’ test performance on factual and applied problems over time?

First, a repeated measures ANOVA was conducted using test presentation mode as the between-subjects variable, and time (test 2, test 3, and test 4) as within-subjects variable for surface style students’ factual problems scores. Second, a repeated measures ANOVA was conducted using test presentation mode as the between-subjects variable and time (test 2, test 3, and test 4) as within-subjects variable for surface style students’ applied problems scores.

3. Does test presentation mode influence the deep style students’ test performance on factual/conceptual and applied problems over time?

First, a repeated measures ANOVA was conducted using test presentation mode as the between-subjects variable, and time (test 2, test 3, and test 4) as within-subjects variable for deep style students’ factual problems scores. Second, a repeated measures ANOVA was conducted using test presentation mode as the between-subjects variable and time (test 2, test 3, and test 4) as within-subjects variable for deep style students’ applied problems scores.

Results

Test 1 was presented in the same test presentation mode (one at a time without backtracking) for all students in the two sections. Test 1 scores were used as a baseline to control any pre-existing differences. The means and standard deviations of students’ overall test 1 scores, factual/conceptual problems scores in test 1, and applied problems scores in test 1 for each section were reported in Table 1.

Table 1. Means and Standard Deviations of Overall, Factual/Conceptual and Applied Scores in Test 1 by Section

Section	Scores	n	M	SD
1	Overall	35	24.83	4.58
	Factual/conceptual	35	18.14	3.26
	Applied	35	6.69	2.18
2	Overall	30	25.40	6.00
	Factual/conceptual	30	19.30	4.21
	Applied	30	6.10	2.20

The one-way MANOVA analyses showed no significant differences between the two sections in overall test 1 scores, factual/conceptual problems scores in test 1, and applied problem scores in test 1. Wilks' $\Lambda = .91$, $F(2, 62) = 3.05$, $p > .05$. However, Levene's test of equality of error variance indicated that the variances of overall test 1 scores and factual/conceptual problems scores in test 1 for each section were significantly different (see Table 2.)

Table 2. Levene's Test of Equality of Error Variances for Test 1 Scores

Scores	F	df 1	df 2	Sig.
Overall	5.59	1	63	.021*
Factual/conceptual	5.13	1	63	.027*
Applied	.009	1	63	.924

* $p < .05$ indicating the rejection of the null hypothesis that the error variance of the dependent variable is equal across groups.

Homogeneity of variances is one of the prerequisite assumptions for ANOVA and MANOVA procedures. If this assumption was violated, it was recommended that a more robust test like Welch's test should be conducted. A Welch's robust test of equality of means was conducted for overall test 1 scores and

factual/conceptual problems scores in test 1. None was significant. For overall test 1 scores, Welch's $F(1, 53.77) = .182$, $p > .05$; for factual/conceptual problems scores in test 1, Welch's $F(1, 54.19) = 1.498$, $p > .05$. Therefore, it was safe to conclude that there were no pre-existing differences in student test performance before they were randomly assigned to the two test-presentation conditions.

Test Presentation Modes and Student Performance in Test 2, Test 3, and Test 4

Research question 1 sought to determine if test presentation modes affect student test performance over time. The means, standard deviations of students' overall test scores, scores on factual problems, and scores on applied problems in test 2, test 3, and test 4 were provided in Table 3.

The repeated measures ANOVA for overall test scores yielded a significant main effect for time $F(2, 126) = 18.19$, $p < .05$. Partial eta squared = .224, indicated relatively large effect size. Follow-up pairwise comparisons revealed that students in both conditions scored significantly higher in test 3 than in test 2 and test 4 ($ps < .05$). The test presentation mode main effect $F(1, 63) = 2.37$, $p > .05$, and time \times presentation mode interaction effect $F(2, 126) = .531$, $p > .05$ were not significant.

To examine whether test presentation mode influences students' factual problems

Table 3. Means and Standard Deviations of Overall Factual, Applied Scores in Test 2, Test 3, and Test 4 by Test Presentation Mode

Mode	Test 2		Test 3		Test 4	
	M	SD	M	SD	M	SD
All at once (n = 35)						
Overall	29.4	4.78	33.06	4.50	29.69	4.86
Factual	15.8	2.61	16.17	2.78	15.57	2.93
Applied	13.6	2.76	16.89	2.19	14.11	2.69
One at a time (n = 30)						
Overall	30.43	4.35	33.83	3.39	31.67	3.92
Factual	16.07	2.23	16.77	2.24	16.60	2.42
Applied	14.37	2.76	17.07	1.66	15.07	2.05

scores, we calculated a second repeated measures ANOVA. No significant effects were found. The main effect for time $F(2, 126) = 1.01, p > .05$, time \times presentation mode interaction $F(2, 126) = .483, p > .05$, and the between-subjects factor (test presentation mode) $F(1, 63) = 1.92, p > .05$ were not significant.

A third repeated measures ANOVA was conducted to examine whether test presentation mode influences students' applied problems scores, the main effect for time was significant $F(2, 126) = 37.86, p < .05$. Partial eta squared = .38, indicated large effect size. Follow-up pairwise comparisons revealed that students in both conditions scored significantly higher in applied problems in test 3 than in test 2 and test 4 ($ps < .05$). The time \times presentation mode interaction $F(2, 126) = .613, p > .05$ and the main effect for test presentation mode $F(1, 63) = 2.22, p > .05$ were not significant.

Surface Style Students' Test Performance in Test 2, Test 3, and Test 4

Research question 2 sought to examine if test presentation modes influence the surface style students' test performance on factual and applied problems over time. The means, standard deviations of surface style students' scores on factual and applied problems in test 2, test 3, and test 4 were provided in Table 4.

The repeated measures ANOVA for surface style students' scores on factual problems revealed a significant main effect for test presentation mode $F(1, 25) = 4.33, p < .05$. Partial eta squared = .15, indicated relatively large effect size. Bonferroni pairwise comparison showed that surface style students in one-at-a-time condition scored significantly higher in factual problems ($p < .05$). The main effect for time $F(2, 50) = 2.36, p > .05$ and time \times presentation mode interaction $F(2, 50) = 1.80, p > .05$ were not significant.

The repeated measures ANOVA for surface style students' scores on applied problems revealed a significant main effect for time $F(2, 50) = 17.8, p < .05$. Partial eta squared = .42, indicated large effect size. Follow-up pairwise comparisons revealed that students in both conditions scored significantly higher in applied problems in test 3 than in test 2 and test 4 ($ps < .05$). No significant time \times presentation mode interaction $F(2, 50) = 1.18, p > .05$, and main effect for test presentation mode $F(1, 25) = 4.16, p > .05$ were found.

Deep Style Students' Test Performance in Test 2, Test 3, and Test 4

Research question 3 sought to examine if test presentation modes influence the deep style students' test performance factual and applied problems over time. The

Table 4. Means and Standard Deviations of Surface Style Students' Factual, Applied Scores in Test 2, Test 3, and Test 4 by Test Presentation Mode

Mode	Test 2		Test 3		Test 4	
	M	SD	M	SD	M	SD
All at once (n = 18)						
Factual	15.78	2.53	16.94	2.21	14.50	3.03
Applied	13.22	2.81	17.17	1.92	13.89	2.68
One at a time (n = 9)						
Factual	16.56	1.67	17.56	2.35	17.33	2.40
Applied	15.11	2.57	17.56	1.01	15.76	1.56

means, standard deviations of deep style students' scores on factual and applied problems in test 2, test 3, and test 4 were provided in Table 5.

The repeated measures ANOVA for deep style students' scores on factual problems revealed non-significant main effect for test presentation mode $F(1, 35) = .151, p > .05$. The main effect for time $F(2, 70) = 1.35, p > .05$ and time \times presentation mode interaction $F(2, 70) = 1.38, p > .05$ were not significant.

The repeated measures ANOVA for deep style students' scores on applied problems revealed a significant main effect for time $F(2, 70) = 14.99, p < .05$. Partial eta squared = .30, indicated large effect size. Follow-up pairwise comparisons revealed that students in both conditions scored significantly higher in applied problems in test 3 than in test 2 and test 4 ($ps < .05$). The main effect for test presentation mode $F(1, 35) = .238, p > .05$, and time \times presentation mode interaction $F(2, 70) = .084, p > .05$ were not significant.

Discussion

Test Presentation Modes and Student Test Performance

The current study evaluated the impact of test presentation modes (all at once vs. one at a time) on four unit tests over the semester. Findings from this study indicated that test presentation modes alone

did not influence student test performance. In addition, test presentation modes alone did not affect student test performance on factual problems or applied problems. These results are consistent with those of Truell, Zhao, and Alexander (2005), who found no significant differences in business students' scores based on the "settable test item exposure control interface format" (p. 31).

Our findings of the relationship between test presentation modes and student test performance further indicated that we need a different mindset toward online learning and assessment. In a paper-pencil test, students usually receive all the test questions at one time in a test booklet. Students can decide the sequence and pace in answering test questions. Similarly, in an online test presented in the all-at-once mode, students can have access to all test questions on their computer screen. They can scroll up and down to decide how they will answer those questions. However, selecting the all-at-once mode did not necessarily influence student test performance. In other words, making online courses emulate the features and functions of a traditional classroom or assessment may not necessarily matter in improving student learning and assessment.

Student Learning Styles, Types of Questions, and Student Test Performance

We found that when the variables of learning styles (surface vs. deep) and types of problems (factual/conceptual vs. applied)

Table 5. Means and Standard Deviations of Deep Style Students' Factual, Applied Scores in Test 2, Test 3, and Test 4 by Test Presentation Mode

Mode	Test 2		Test 3		Test 4	
	M	SD	M	SD	M	SD
All at once (n = 16)						
Factual	15.75	2.84	15.31	3.24	16.75	2.49
Applied	14.00	2.80	16.50	2.53	14.31	2.85
One at a time (n = 21)						
Factual	15.86	2.43	16.43	2.16	16.29	2.41
Applied	14.05	2.84	16.86	1.85	14.76	2.19

were added, test presentation mode had a positive effect on student test performance over time. Surface style students scored significantly higher on factual/conceptual problems in the one-at-a-time test presentation mode than in the all-at-once mode. Factual/conceptual questions measure the lower level skills in the cognitive domain (i.e., knowledge and comprehension in Bloom's taxonomy). The one-at-a-time test presentation mode helps surface style students focus on the targeted question, which may lead to success in recalling what they have memorized. This interpretation of our results is consistent with the recent study of Steinmayr, Ziegler, and Träuble (2010), in which they explained the interplay between sustained attention, intelligence, and school performance in a non-clinical sample.

Recommendations and Limitations

How should test presentation modes in Blackboard® be selected for our online tests? This study suggests that there is no magic test presentation option in Blackboard® for boosting students' test performance. However, surface style students may benefit from the one-at-a-time test presentation mode over time. Understanding the students will help online course instructors and online test designers

choose an optimal test presentation mode for their students. In addition, technical issues and test security should also be taken into consideration when determining the test presentation mode for an online test. The one-at-a-time mode usually requires more server resources, which could lead to server overload. The all-at-once mode will make it hard to protect the content of an online test. These criteria may also be applicable to other interactive student response system, such as the Classroom Performance System™ (CPS or clickers), which are widely used in the public schools.

A limitation to the present study was that students were not randomly selected. In addition, this study could use a reliable and valid learning style inventory in addition to self-reported written responses. Therefore, it would be prudent to conduct additional studies to consolidate these conclusions.

Acknowledgements

This project was partially supported by a 2009-2010 Regular Research Grant from Committee on Institutional Studies and Research (CISR), Murray State University. The author is grateful to George Patmor, and the anonymous reviewers for their insightful contributions to this paper.

References

- Barrada, J., Olea, J., Ponsoda, V., & Abad, F. (2008). Incorporating randomness in the Fisher information for improving item-exposure control in CATs. *British Journal of Mathematical & Statistical Psychology*, 61(2), 493-513.
- Blackboard® [Software]. Washington, DC: Blackboard Inc.
- Bloom, B., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *A taxonomy of educational objectives, handbook 1: The cognitive domain*. New York: David McKay.
- Coffield, F. J., Moseley, D. V., Hall, E., & Ecclestone, K. (2004). *Learning styles and pedagogy for post-16 learning: A systematic and critical review*. London: Learning and Skills Research Centre.
- Doong, S. (2009). A knowledge-based approach for item exposure control in computerized adaptive testing. *Journal of Educational & Behavioral Statistics*, 34(4), 530-558.
- Dunn, R. (1987). Research on instructional environments: Implications for student achievement and attitudes. *Professional School Psychology*, 2, 43-52.
- Dunning, D. (2008). Learning style. In N. J. Salkind (Ed.), *Encyclopedia of educational psychology*. Thousand Oaks, CA: Sage.
- Finkelman, M., Nering, M., & Roussos, L. (2009). A conditional exposure control method for multidimensional adaptive testing. *Journal of Educational Measurement*, 46(1), 84-103.
- Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, 23(3), 187-194.
- Pastor, D., Dodd, B., & Chang, H. (2002). A comparison of item selection techniques and exposure control mechanisms in cats using the generalized partial credit model. *Applied Psychological Measurement*, 26(2), 147-163.
- Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(4), 311-327.
- Snow, R. E., Corno, L., & Jackson, D. (1996). Individual differences in affective and conative functions. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 243-310). New York: Simon & Schuster.
- Stahl, S. A. (2002). Difference strokes for different folks? In L. Abbeduto (Ed.), *Taking sides: Clashing on controversial issues in educational psychology* (pp. 98-107). Guilford, CT: McGraw-Hill/Duskin.
- Steinmayr, R., Ziegler, M., & Träuble, B. (2010). Do intelligence and sustained attention interact in predicting academic achievement? *Learning & Individual Differences*, 20(1), 14-18.
- Stocking, M. L., Ward, W. C., & Potenza, M. T. (1998). Simulating the use of disclosed items in computerized adaptive testing. *Journal of Educational Measurement*, 35(1), 48-68.
- Test options. (2006). In *Blackboard® Instructor Manual*. Retrieved from http://library.blackboard.com/ref/8b752651-5c02-4644-9340-69330d559013/test_options.htm
- Truell, A. D., Zhao, J. J., & Alexander, M. W. (2005). The impact of settable test item exposure control interface format on postsecondary business student test performance. *Journal of Career and Technical Education*, 22 (1), 31-41.
- Van der Linden, W., & Veldkamp, B. (2007). Conditional item-exposure control in adaptive testing using item-ineligibility probabilities. *Journal of Educational and Behavioral Statistics*, 32(4), 398-418.

Veldkamp, B., & van der Linden, W. (2008). Implementing Simpson-Hetter item-exposure control in a shadow-test approach to constrained adaptive testing. *International Journal of Testing*, 8(3), 272-289.

Winne, P. H. (1979). Experiments relating teachers' use of higher cognitive questions to student achievement. *Review of Educational Research*, 49, 13-50.

Woolfolk, A. (2008). *Educational psychology*. Boston, MA: Pearson.

Yuejin Xu is Assistant Professor, College of Education, Murray State University, Murray, Kentucky