Running head: RELATIONSHIP BETWEEN STUDENTS' FORMATIVE EXAM-TAKING

Relationship Between Students' Formative Exam-Taking Strategies and Summative Exam Scores

Beverly M. Klecker

Morehead State University

b.klecker@moreheadstate.edu

Richard L. Klecker

University of Kentucky

Paper presented at the annual meeting of

The Mid-South Educational Research Association

November 5, 2010

Columbus, OH

Abstract

This study explored graduate students' formative exam-taking strategies in an online classroom. Research questions were: (1) What formative test-taking strategies did students use? (2) Were the strategies related to the students' final exam scores? Students (N=50) in a semester-long measurement class took weekly formative 20-item, multiple-choice exams. Students' options were: (1) using a printable copy to guide their reading then entering their answers in an identical answer sheet for scoring, and/or (2) using only the answer sheet interactively until reaching mastery. Data were the number of times students' entered answers to reach mastery and their final exam scores. Correlation between the number of times students took the exams (summed across 16 weeks) and their final exam scores was rs =0.68.

Relationship Between Students' Formative Exam-Taking Strategies and Summative Exam

The purpose of this study was to examine the patterns of graduate students' formative exam-taking strategies and their scores on the summative final exam. The study adds to the growing body of research on formative assessment (e.g., Brookhart, 2004; Chappius, Chappius, and Stiggins, 2009; Guskey, 2010; Klecker, 2003; Linn & Miller, 2005; Ricketts, & Wilks, 2002) and assessment in online classrooms (Beebe, Vonderwell, & Boboc, 2010; Benson, 2003; Klecker, 2007). Additionally, the study explored the use of spacing presentation of material and using test-taking as a learning strategy. Rohrer and Pashler (2010) presented a discussion of learning through testing in which they cite numerous studies across years, They stated, "…More recent studies, however, have shown that a combination of study and tests is more effective than spending the same amount of time reviewing the material in some other way, such as rereading it…" (p. 406).

Background of the Study

Formative assessment was used in this study as "…assessment for learning…" (Stiggins & Chappuis, 20008, p. 43). Mastery learning (Guskey, 2005; Klecker & Chapman, 2008) was modeled by permitting students to re-take the exams until they reached mastery (20/20). Findings from an earlier quasi-experimental study (Klecker, 2007) supported the effectiveness of using weekly formative exams for learning. Additionally, student course evaluations consistently supported the use of the formative exams.

Structure of the class. Materials in the online class were presented in sixteen weekly sections. Reading assignment in the textbook, supplemental materials in the form of Power Points and External Links, the weekly exam (in two formats), and the weekly discussion board

were posted on Blackboard each week on Monday at 7:00 AM. The exam and the discussion board were completed each week by Sunday 11:55 PM.

Formative exams were available each week in two formats (1) printable; and (2) an "answer sheet," identical to the printable version in Bb to be used to submit answers for immediate feedback. The students received feedback in the form of the number of questions that they had answered correctly and the items that were not answered correctly. Correct answers were never given to the students until they answered the question correctly.

When the 20 multiple-choice questions were entered into Bb answer sheet, the items were consistently presented in the same order, but the alternatives were "randomized" by Bb each time the student exited/re-entered the Exam. This was done to avoid the "AAAA, BBBB, CCCC, DDDD" strategy for the four-option questions. Thus, students had to look at the words in the alternatives rather than just the letter.

Student strategies for formative exams. Students could: (1) print test and mark the answers while reading the assignment then enter answers on Bb; and/or (2) use iterative process by entering answers directly on Bb. Option 1 was the suggested strategy in the Syllabus:

> Print the printable copy and use it to guide your reading in the assigned chapter. Then enter the Answer Sheet and submit your answers for scoring. You may enter/exit the Answer Sheet until you have reached mastery of 20/20 by Sunday 11:55 PM.

The Final Exam was available Monday-Friday of Finals Week . The final exam—testing the same concepts with different questions-- was in printable and answer sheet form; *students could enter answers only once on the final exam*.

Grading for the class. This class was taught for mastery and used criterion-referenced grading. The scores from the 16 formative exams comprised 20% of the student's final grade; the

score on the final exam comprised 30%. Additionally, the grade for the class was comprised of grades on weekly Discussion Boards (25%) and grade on a test-evaluation paper using the *Mental Measurement Yearbook* (Buros Institute, 2010) as a resource (35%).

Limitation of the Study

The scope and limited non-random sample in this study limits generalizability of the results. Additionally, it should be noted that the weekly Discussion Boards consisted of a brief open-open response question that required the student to collect, evaluate, and synthesize information on the week's topic. Students were required to post a response to one other student's reply to the question. The weekly open-ended questions on the same concepts measured by the multiple-choice likely influenced students' achievement on the final exam (Rohrer & Pashler, 2010).

Research Questions

The Research Questions for this exploratory research study were: (1) What formative test-taking strategies did students use? (2) Were the strategies used related to scores on the final exam?

Method

Participants

Participants for the study were 50 master's-level students at a public state university in Kentucky. The students self-enrolled in one of two identical Blackboard sections of the course in measurement principles and techniques during the 2009-2010 school year. The sample was non-random. The students in the class were in varied graduate programs in the College of Education. The course is required for several master's degrees including; master's in school counseling,

master's of arts in teaching (MAT)--a graduate initial teacher certification degree, master's in educational leadership.

Instrumentation

Weekly formative exams. Some of the items for the formative exams were selected from the textbook's item bank, many others were created by the instructor in the class. Since the class is taught for mastery and grading is criterion-referenced, questions were first selected for content validity. Bloom's (1956) cognitive levels were used as estimates of item difficulty. Since concepts, vocabulary, and calculations were new to students, the 20-item formative exams typically included eight knowledge-level questions, seven comprehension-level questions, and five application-level questions. (Students frequently e-mailed questions about the application-level questions the week after the test had been submitted.)

Final exam. The 60-multiple-choice-item final exam contained three or four questions from each chapter. The distribution of cognitive levels (Bloom, 1956) was 20 knowledge level, 15 comprehension-level, and 35 application-level. Students receive the printable copy and answer sheet Monday of finals week and must submit their answers only once by Friday 11:55 PM. Feedback given to students is the number of items correct and the items that were incorrect. Correct answers are not given.

Procedure

Approval for the study was obtained from the university's Institutional Review Board. After students' grades had been submitted to the Registrar's Office at the end of the semester, the data were downloaded from Blackboard. .

Over time, students typically adopted two strategies.

- First strategy: Students printed the test and marked the answers as they read the assigned reading in their textbooks, PowerPoints, and on External Links. After selecting their answers on the "printable copy," they then used the "answer sheet" to enter their answers. They were able to exit/re-enter/re-submit until they reached mastery.
- Second strategy: Students did not use the printable copy. They went directly to the "answer sheet" and used iterations of enter/marking/scoring/exiting until reaching mastery of 20/20.

Data Analysis

Data were downloaded into an Excel spreadsheet and uploaded into SAS® software V. 9 for Windows. Descriptive statistics and correlations were computed. Data analysis included descriptive statistics for: (1) the number iterations for mastery (2) students' scores on the final exam.

Results

Spearman correlation for the data revealed that the summed number of iterations and the final exam scores were statistically significantly related,

$rs = 0.68$, $n = 50$, $p.<.001$, two tailed.

Table 1 presents the topic for each week, the maximum and minimum number of times students took the exam for the week and the mode for each week.

[Insert Table 1 about here]

The data for each week look similar with a minimum of 1 iteration and a maximum of 6 or 7. The modal value for each week, with the exception of Week 11 was 4. Week 11 had a minimum of 1, a maximum of 7 and a mode of 2. The topic for Week 11 was "Standards for

Selecting and Evaluating Standardized Tests," a chapter that was partially a summary of material presented in previous chapters.

[Insert Table 2 about here]

Table 2 presents the frequencies and percentages of the number of times that students took each weekly exam. Boldface type indicated the modal response for each week. The spread of students' strategies across the number of iterations indicate that there were some students each week who reached mastery after one submission of their answers and there were students who required more than four (the number of alternatives or options) submissions to reach mastery.

[Insert Table 3 about here]

Spearman correlation for the data revealed that the summed number of iterations and the final exam scores were statistically significantly related, $rs = 0.68$, $n = 50$, $p.<.001$, two tailed.

[Insert Table 4 about here]

Table 4 presents disaggregated data by week and looked for correlations with the score on the final exam. This Table probably best describes the effect of restriction of range on correlation coefficients. The p-values are included adhering to convention, but are meaningless and are a function of the size of the non-random sample. (N=50).

Conclusions

The teacher/researcher for this study feels strongly that graduate classes---especially in research methods and measurement principles and techniques (the class in this study) should be taught for mastery. That is, the student should have every opportunity to learn the material through formative evaluation that he or she is allowed to complete for a mastery formative grade. In the student evaluation of this course, the students rate the formative exams (30% of final grade) as their favorite part of the class. Many online instructors are still concerned that "someone else" will take a high-stakes final exam for the student. This researcher's view is that

instructors should be asking different questions and have other concerns---namely about the *quality of the exam* and the student's motivation to learn through exams.

References

Beebe, R., Vonderwell, S., & Boboc, M. (2010). Emerging patterns in transferring assessment practices from F2f to online environments. *Electronic Journal of e-Learning, 8*, (1), 1-12.

Benson, A.D. (2003). Assessing participant leaning in online environments. *New Directions for Adult and Continuing Education, 100*, 69-77.

Brookhart, S.M. (2004). *Grading*. Upper Saddle River, NJ: Pearson, Merrill, Prentice-Hall.

Buros Institute of Mental Measurements (2010). *Mental Measurement Yearbooks* Retrieved from the Buros Institute's *Test Reviews online* website: http://www.unl.edu/buros/bimm/html/subburos.html

Chappuis, S., Chappuis, J., Stiggins, R. (November, 2009). The quest for quality. *Educational Leadership, 67*, (3), 14-19.

Guskey, T. R. (2005). *Formative classroom assessment and Benjamin S. Bloom: Theory, reseach, and implications*. Paper presented at the Annual Meeting of the American Educational Research Association. Montreal, Canada. (ERIC Document Reproduction Service No. ED490412)

Guskey, T.R. (2010). Formative assessment: The contributions of Benjamin S. Bloom, In C.J. Cizek and H.L. Andrade (Eds.), *Handbook of formative assessment* (pp. 106-124), New York, Routledge.

Klecker, B. (2007) The impact of formative feedback on student learning in an online classroom. *Journal of Instructional Psychology, 34.* (3),161-165 (ERIC Document Reproduction Service No. (EJ778785)

Klecker, B. & Chapman, A. (2008) *Advocating the Implementation of Mastery Learning in Higher Education to Increase Student Learning and Retention* Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN (ERIC Document Reproduction Service No. ED503410)

Linn, R. L., & Miller, M.D. (2005). *Measurement and assessment in teaching* (9th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Ricketts, C., & Wilks, S. J. (2002). Improving student performance through computer-based assessment: Insights from recent research. *Assessment & Evaluation in Higher Education, 27,* (5), 475-479.

Rohrer, D. & Pashler, H. (June/July 2010). Recent research on human learning challenges and conventional instructional strategies. *Educational Researcher, 39,* (5), 406-412.

Stiggins, R., & Chappuis, J. (2008). Enhancing student learning. *District Administration, 44*, (1), 42-44.

Table 1. Weekly Topics by the Number of Times Students' Took Formative Exam

Week and Topic	N	Minimum	Maximum	Mode
1. Fundamentals of Assessment	50	1	7	4
2. Legal, Social, and Ethical Implications of Testing	50	1	6	4
3. Planning the Teacher-Made Test	50	1	7	4
4 Principles of Test Construction: Selection Type Items	50	1	6	4
5 Principles of Test Construction: Supply Type Items	50	1	6	4
6 Evaluating Items: Item Analysis and Interpretations	50	1	5	4
7 Summarizing and Interpreting Measurements: Descriptive Statistics	50	1	9	4
8 Summarizing and Interpreting Measurements	50	1	7	4
9 The Reliability of Measurements	50	1	6	4
10 The Validity of Inferences Obtained from Tests and Measurements	50	1	9	4
11 Standards for Selecting and Evaluating Standardized Tests	50	1	5	2
12 The Nature and Measurement of Intelligence and Aptitudes	50	1	8	4
13 The Measurement of Aptitudes	50	1	9	4
14 Standardized Measures of Academic Achievement (CATS, CTBS, NAEP)	50	1	7	4
15 The Measurement of Interests, Attitudes, and Values	50	1	6	4

16 The Measurement of Personality Traits	50	1	7	4

Table 2. Frequencies and Percentages Number of Times Students' Took Weekly Exams

Week	Number of Times to Mastery	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Week 1	1	7	14.00%	7	14.00
	2	7	14.00%	14	28.00
	3	11	22.00%	25	50.00
	4	**17**	**34.00**%	**42**	**84.00**
	5	6	12.00%	48	96.00
	6	1	2.00%	49	98.00
	7	1	2.00%	50	100.00
Week 2	1	6	12.00%	6	12.00
	2	5	10.00%	11	22.00
	3	10	20.00%	21	42.00
	4	**20**	**40.00**%	**41**	**82.00**

	5	7	14.00%	48	96.00
	6	2	4.00%	50	100.00
Week 3	1	6	12.00%	6	12.00
	2	10	20.00%	16	32.00
	3	9	18.00%	25	50.00
	4	**18**	**36.00**%	**43**	**86.00**
	5	5	10.00%	48	96.00
	6	1	2.00%	49	98.00
	7	1	2.00%	50	100.00
Week 4	1	5	10.00%	5	10.00
	2	6	12.00%	11	22.00
	3	15	30.00%	26	52.00
	4	**18**	**36.00**%	**44**	**88.00**
	5	5	10.00%	49	98.00
	6	1	2.00%	50	100.00

(Table continued next page)

Table 2. Frequencies and Percentages Number of Times Students' Took Weekly Exams (Continued)

	Number of Times to Mastery	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Week 5	1	3	6.00%	3	6.00
	2	5	10.00%	8	16.00
	3	15	30.00%	23	46.00
	4	**23**	**46.00**%	**46**	**92.00**
	5	3	6.00%	49	98.00
	6	1	2.00%	50	100.00
Week 6	1	4	8.00%	4	8.00
	2	6	12.00%	10	20.00
	3	6	12.00%	16	32.00
	4	**31**	**62.00**%	**47**	**94.00**
	5	3	6.00%	50	100.00
Week 7	1	1	2.00%	1	2.00
	2	6	12.00%	7	14.00

	3	9	18.00%	16	32.00
	4	**30**	**60.00**%	**46**	**92.00**
	5	2	4.00%	48	96.00
	6	1	2.00%	49	98.00
	9	1	2.00%	50	100.00
Week 8	1	3	6.00%	3	6.00
	2	7	14.00%	10	20.00
	3	12	24.00%	22	44.00
	4	**22**	**44.00**%	**44**	**88.00**
	5	4	8.00%	48	96.00
	6	1	2.00%	49	98.00
	7	1	2.00%	50	100.00

(Table continued on next page)

Table 2. Frequencies and Percentages Number of Times Students' Took Weekly Exams (Continued)

	Number of Times to Mastery	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Week 9	1	5	10.00%	5	10.00
	2	8	16.00%	13	26.00
	3	4	8.00%	17	34.00
	4	**26**	**52.00**%	**43**	**86.00**
	5	4	8.00%	47	94.00
	6	3	6.00%	50	100.00
Week 10	1	4	8.00%	4	8.00
	2	8	16.00%	12	24.00
	3	10	20.00%	22	44.00
	4	**21**	**42.00**%	**43**	**86.00**
	5	4	8.00%	47	94.00
	6	2	4.00%	49	98.00
	9	1	2.00%	50	100.00

Week 11	1	11	22.00%	11	22.00
	2	**29**	**58.00**%	**40**	**80.00**
	3	6	12.00%	46	92.00
	4	3	6.00%	49	98.00
	5	1	2.00%	50	100.00
Week 12	1	4	8.00%	4	8.00
	2	2	4.00%	6	12.00
	3	4	8.00%	10	20.00
	4	**23**	**46.00**%	**33**	**66.00**
	5	11	22.00%	44	88.00
	6	2	4.00%	46	92.00
	7	2	4.00%	48	96.00
	8	2	4.00%	50	100.00

(Table continued next page)

Table 2. Frequencies and Percentages Number of Times Students' Took Weekly Exams (Continued)

	Number of Times to Mastery	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Week 13	1	2	4.00%	2	4.00
	2	6	12.00%	8	16.00
	3	3	6.00%	11	22.00
	4	**27**	**54.00**%	**38**	**76.00**
	5	8	16.00%	46	92.00
	7	1	2.00%	47	94.00
	8	2	4.00%	49	98.00
	9	1	2.00%	50	100.00
Week 14	1	5	10.00%	5	10.00
	2	11	22.00%	16	32.00
	3	11	22.00%	27	54.00
	4	**17**	**34.00**%	**44**	**88.00**
	5	5	10.00%	49	98.00
	7	1	2.00%	50	100.00

Week 15	1	3	6.00%	3	6.00
	2	2	4.00%	5	10.00
	3	11	22.00%	16	32.00
	4	**28**	**56.00**%	**44**	**88.00**
	5	4	8.00%	48	96.00
	6	2	4.00%	50	100.00
Week 16	1	2	4.00%	2	4.00
	2	4	8.00%	6	12.00
	3	4	8.00%	10	20.00
	4	**28**	**56.00**%	**38**	**76.00**
	5	8	16.00%	46	92.00
	6	3	6.00%	49	98.00
	7	1	2.00%	50	100.00

Table 3. Sum of Number of Times Student Took Formative Exam by Final Exam Score

Student	Sum of Number of Times Test Taken (16 Exams)	Final Exam Score (Percent correct)	Student	Sum of Number of Times Test Taken (16 Exams)	Final Exam Score (Percent correct)
1	27	90.00%	**26**	61	100.00%
2	27	61.67%	**27**	61	98.33%
3	29	98.30%	**28**	61	100.00%
4	36	98.33%	**29**	62	98.33%
5	36	100.00%	**30**	62	100.00%
6	39	96.67%	**31**	62	100.00%
7	40	100.00%	**32**	62	98.33%
8	41	98.33%	**33**	63	98.33%
9	43	98.33%	**34**	63	100.00%
10	45	95.00%	**35**	63	100.00%
11	47	81.00%	**36**	63	100.00%
12	48	98.33%	**37**	64	90.00%
13	48	96.67%	**38**	64	100.00%
14	50	96.67%	**39**	64	98.33%
15	51	81.00%	**40**	65	98.33%
16	53	100.00%	**41**	65	96.97%
17	53	98.33%	**42**	65	86.67%
18	54	100.00%	**43**	65	91.67%
19	56	100.00%	**44**	66	81.67%

20	56	98.33%	**45**	66	96.67%
21	57	83.33%	**46**	66	98.33%
22	57	98.33%	**47**	69	98.33%
23	57	100.00%	**48**	69	100.00%
24	59	100.00%	**49**	69	100.00%
25	59	100.00%	**50**	70	86.67%

Table 4. Spearman Correlations of Test-Taking Frequencies for Weekly Exams with Final Exam Score

Week and Topic	**Correlation with Final Exam Score**
1. Fundamentals of Assessment	0.107 *p.*<0.457
2. Legal, Social, and Ethical Implications of Testing	0.017 *p.*<0.901
3. Planning the Teacher-Made Test	0.071 *p.*<0.625
4 Principles of Test Construction: True-False, Multiple-Choice and Matching Items.	0.105 *p.*<0.469
5 Principles of Test Construction: Completion, Open-Response Items, and Essay Tests.	0.152 *p.*<0.291
6 Evaluating Items: Item Analysis and Interpretations	0.078 *p.*<0.589
7 Summarizing and Interpreting Measurements: Descriptive Statistics	0.129 *p.*<0.374
8 Summarizing and Interpreting Measurements: Norms and Score Interpretation	-0.118 *p.*<0.416
9 The Reliability of Measurements	0.0605 *p.*<0.677
10 The Validity of Inferences Obtained from Tests and Measurements	0.181 *p.*<0.208
11 Standards for Selecting and Evaluating Standardized Tests	-0.065 *p.*<0.653

12 The Nature and Measurement of Intelligence and Aptitudes	0.290	*p.*<0.041
13 The Measurement of Aptitudes	0.163	*p.*<0.259
14 Standardized Measures of Academic Achievement (CATS, CTBS, NAEP)	-0.036	*p.*<0.803
15 The Measurement of Interests, Attitudes, and Values	0.065	*p.*<0.652
16 The Measurement of Personality Traits	0.291	*p.*<0.04

Table 4. Spearman Correlations of Test-Taking Frequencies for Weekly Exams with Final Exam Score

Week and Topic	Correlation with Final Exam Score
1. Fundamentals of Assessment	0.107 *p.*<0.457
2. Legal, Social, and Ethical Implications of Testing	0.017 *p.*<0.901
3. Planning the Teacher-Made Test	0.071 *p.*<0.625
4 Principles of Test Construction: True-False, Multiple-Choice and Matching Items.	0.105 *p.*<0.469
5 Principles of Test Construction: Completion, Open-Response Items, and Essay Tests.	0.152 *p.*<0.291
6 Evaluating Items: Item Analysis and Interpretations	0.078 *p.*<0.589
7 Summarizing and Interpreting Measurements: Descriptive Statistics	0.129 *p.*<0.374
8 Summarizing and Interpreting Measurements: Norms and Score Interpretation	-0.118 *p.*<0.416
9 The Reliability of Measurements	0.0605 *p.*<0.677
10 The Validity of Inferences Obtained from Tests and Measurements	0.181 *p.*<0.208
11 Standards for Selecting and Evaluating Standardized Tests	-0.065 *p.*<0.653
12 The Nature and Measurement of Intelligence and Aptitudes	0.290 *p.*<0.041
13 The Measurement of Aptitudes	0.163 *p.*<0.259
14 Standardized Measures of Academic Achievement (CATS, CTBS, NAEP)	-0.036 *p.*<0.803
15 The Measurement of Interests, Attitudes, and Values	0.065 *p.*<0.652
16 The Measurement of Personality Traits	0.291 *p.*<0.04