Institute for International Economic Policy Working Paper Series
Elliott School of International Affairs
The George Washington University

Test Format and Calculator Use in the Testing of Basic Math Skills for Principles of Economics: Experimental Evidence

IIEP-WP-2020-20

Irene R. Foster
George Washington University

Melanie Allwine Fennell
Randolph-Macon College

October 2020

Institute for International Economic Policy
1957 E St. NW, Suite 502
Voice: (202) 994-5320
Fax: (202) 994-5477
Email: **iiep@gwu.edu**
Web: **iiep.gwu.edu**

TEST FORMAT AND CALCULATOR USE IN THE TESTING OF BASIC MATH SKILLS FOR PRINCIPLES OF ECONOMICS: EXPERIMENTAL EVIDENCE

Melanie Allwine Fennell
Randolph-Macon College
114 College Ave.
Ashland, VA 23005
melaniefennell@rmc.edu
704-497-0469

Irene R. Foster
George Washington University
2115 G St. NW Suite 340
Washington, DC 20052

Institutional Review Board permission was granted to conduct this research at George Washington University (IRB #081330)

ABSTRACT

Results from an experiment in Fall 2013 of 902 incoming students at this university are reported. In this experiment, after students were given a basic math assessment to ensure they had the necessary math skills to take a principles of economics course, they were randomly allocated to a treatment or control group to test if there was a significant impact of test format, calculator use, and calculator type on students' scores. The interaction of calculator use/type and test format was also tested. The results from this experiment suggest that each treatment had a significant positive impact on students' assessment scores, with much variation depending on the type of question asked and the level of performance.

Keywords: Economic Education, Teaching Economics, Math Assessment, Microeconomics, Calculator Use, Test Format

JEL codes: A22, C23

INTRODUCTION

Test format matters when assessing college students' mathematics and economic knowledge (Frederiksen, 1984; Bridgeman, 1992; Katz, Bennett, and Burger, 2000; Chan and Kennedy, 2006). Bridgeman (1992) finds that students perform better on certain GRE math problems when given multiple choice in comparison to an open ended framework. Chan and Kennedy (2002) find that students perform better on multiple choice questions than "equivalent" critical response questions in principles of macroeconomics.

Calculator use can have a significant impact on college students' ability to correctly answer basic math questions (Boyle and Farreras, 2015; Koop, 1982). Boyle and Farreras (2015) find that having a calculator has a significant positive effect on the proportion of students that correctly answer certain questions when evaluating college students on basic math skills. Schwarz et al.

(2002) finds that the type of calculator that students use matters when assessing algebra 1 knowledge in high school students.

Basic math understanding is critical for students to do well in a principles of economics course (Mallik & Lodewijks, 2010; Owen, 2012; Allgood et al., 2015; Durden & Ellis, 1995; Hafer & Hafer, 2002; Mallik & Shankar, 2016; Milkman et al., 1995). Research has found that students' performance on a basic math assessment or quiz has a positive correlation with performance in principles of economics (Ballard & Johnson, 2004; Fennell & Foster, 2019; Schuhmann et al., 2005; Douglas & Sulock, 1995; Hafer & Hafer, 2002). Ballard and Johnson (2004) and Schuhmann et al. (2005) use a multiple choice format for their assessment. Fennell and Foster (2019) use an open-ended format and do not allow for the use of a calculator for their assessment.

Economics faculty at this university administer a basic math assessment (BMA) to all principles of economics students. The BMA is 20 simple questions that cover only those skills required for a principles class: percentage change, ratios, proportions, fractions and decimals, order of operations, place value, the area of a triangle, simple exponents, and the graph of a straight line. A student is deemed to have mastered these basic skills if they can correctly answer 16 or more of the 20 questions, allowing students a few simple mistakes. The assessment includes only open-ended questions and does not allow the use of a calculator, thereby following the format used in Fennell and Foster (2019). Given that test format and calculator use have a significant impact on a student's ability to correctly answer basic math problems, we hypothesize that the format of the assessment given to incoming economics students will impact their assessment scores.

In this paper, we analyze the impact of test format and calculator use on students' ability to answer basic math questions utilized in principles of economics courses and the interactive effect of these two by running a randomized experiment in the university's principles of

microeconomics course. The experimental design was to give students an additional assessment (referred to as a questionnaire) after the BMA and randomly assign students to the use of one of three types of calculators (basic, graphing, and own calculator). Then, randomly assign students a test format (multiple choice or open-ended) to analyze not only the impact of calculator use/type and test format on students math skills necessary to do well in a principles of economics course but also the impact of the interactive effect of test format and calculator use/type. Once data was collected, this was tested for using a difference in difference framework empirically.

Our results suggest that calculator use/type and test format have a significant positive impact on students overall scores on a basic math assessment. Use of a calculator increased the average student's score by one half of a point, and multiple-choice framework increased their score by one point out of 10 points total. Similar to the current literature, we find that the effect of calculator use and test format differs depending on the type of question asked. Multiple choice has a significant effect on the likelihood of students answering the question correctly on most problems and therefore has a larger impact on total score. Calculator use has a significant effect on a few problems, specifically algebraic word problems with a high level of difficulty. In this type of problem, there is no evidence that the type of calculator has any effect. These effects are larger for lower-performing students.

The remainder of this paper is organized as follows: the second section gives a literature review and description of the BMA, the third section describes the methodology, the fourth section presents results, and the final section concludes.

LITERATURE REVIEW

Many researchers have found that math skills are important to students' performance in economics courses (Mallik & Lodewijks, 2010; Jensen & Owen, 2003; Ballard & Johnson, 2004;

Hoag & Benedict, 2010; Douglas & Sulock, 1995; Gallo & Johnson, 2008; Schuhmann et al., 2005; Durden & Ellis, 1995; Hafer & Hafer, 2002; Mallik & Shankar, 2016; Milkman et al., 1995; Kroncke & Smyth, 2003). These studies measure math skills in a range of ways, including SAT and ACT math scores, performance in high school math classes, whether the student has taken calculus, and enrollment in a remedial math class. Research has also found that principles of economics courses improve students' math skills (Mendez-Carbajo, et al., 2018; Smyth and Kroncke, 2005; Pinter, 2014).

Schumann et al. (2005), Ballard and Johnson (2004), and Fennell and Foster (2019) find that, in addition to good standardized test scores, a student's performance on a basic mathematics assessment has a statistically significant positive correlation with performance in an economics course. Schuhmann et al. (2005) find that students that improved their math scores on a basic math assessment had a higher probability of scoring more questions correctly on a survey of principles of microeconomics material. Ballard and Johnson (2004) find that taking a college calculus course, performance on a basic math assessment, and SAT scores have a positive effect on performance in principles of economics. The authors conclude that quantitative skills are multidimensional, and therefore, many measures of mathematics skills will prove to be significant determinants of economics understanding. Fennell and Foster (2019) find that a basic math assessment has a statistically significant positive correlation with performance in principles of microeconomics and that additional attempts on the assessment predict better performance. The format of the assessment given in Schuhmann et al. (2005) and Ballard and Johnson (2004) is multiple choice, whereas the format of the assessment in Fennell and Foster (2019) is open-ended. Calculators are also not allowed in the assessment in Fennell and Foster (2019).

Research on the impact of calculator use on college age students' ability to correctly answer basic math questions has found that students do better on basic math questions when they are allowed to use a calculator, but this effect differs depending on the type of question that is asked. In a basic arithmetic course at the community college level, Koop (1982) finds that, when students are randomly assigned to receive a calculator in a community college arithmetic course, they do better. Bridgeman, Harvey, and Braswell (1995) find that the benefit/cost of using a calculator on the SAT depends on the type of SAT math question asked. A student is more likely to be able to answer multiplication, division, and exponent problems when using a calculator. They find no effect on algebraic word problems or problems about finding the area of a triangle. Boyle and Farreras (2015) find that college-level psychology students that are given a calculator on a simple math assessment performed better on basic multiplication, addition, and subtraction questions with no difference in performance on algebraic word problems. Scheuneman et al. (2002) find that students with better math skills are more likely to use a calculator and therefore score better on the SAT math. Schwarz et al. (2002) finds that the positive impact of calculator use for algebra I questions is larger for students with weaker math skills, but that this benefit is for low-level computation questions. All of these assessments used multiple choice questions.

Research on the differential impact of the type of calculator used on students' ability to correctly answer basic math questions for college-age students has found mixed results. Bridgeman and Potenza (1998) compare scores on SAT math questions when students use their own calculator versus an on-screen calculator. They find no impact of using the provided calculator versus students' own calculators. Schwarz et al. (2002) finds that students who use their own graphing calculators to answer algebra I assessment questions perform better than students who do not use any calculator.

Research on the impact of different test frameworks on college students' ability to correctly answer basic math questions suggests mixed results. Katz, Bennett, and Burger (2000) find that test format has little to no effect on basic arithmetic problems but a large effect on algebraic problems (especially word problems that increase in difficulty). Authors suggest that this is due to nontraditional methods, such as plugging in solutions, as well as students getting an answer through their own calculations but not finding their answer in the multiple choice options, which allows them to identify the error in their calculations. Bridgeman (1992) uses GRE questions to perform an experiment to determine the effect of test format on ability to correctly answer GRE questions. He finds that, for complicated multiplication, division, and percentage problems, there is a significant positive effect that is especially high for low performers. He also suggests that format effects are larger when multiple choice options do not reflect the errors that students actually make. There is no effect on performance on algebraic word problems. Frederiksen (1984) finds that, when critical thinking problems are converted to multiple choice problems, the correlation between multiple choice and constructed response questions are low and that performance is much higher on multiple choice problems with less critical thinking occurring.

Research on the impact of different test frameworks on students' ability to correctly answer economics questions finds that students perform better overall on the multiple choice questions than critical response questions (Chan & Kennedy, 2002). To our knowledge, no one has studied the interactive effect of test framework and calculator use on performance on basic math questions or in economics. A different strain of research that is parallel to the research above focuses on the difference in skills tested across different frameworks and calculator use (Becker & Johnston; 1999, Rebeck & Asarta, 2012; Buckles & Siegfried, 2006; O'Neil & Brown, 2009). This paper assumes that questions are constructed to achieve "equivalent" skill assessment goals.

Basic Math Assessment

Due to faculty members' observations that students were weak in mathematics and that this inhibited their learning of economics, in the fall of 2010, faculty began giving a basic math assessment (BMA) to all incoming principles of economics students. The assessment counts towards 10% of each student's grade, thereby demonstrating the importance of math skills in the understanding of principles courses. The framework of the assessment is open-ended, and no calculators are allowed. Students are provided a box for their answer, and answers are graded as correct or incorrect. Calculator use and test framework were highly debated within the faculty and this methodology was decided upon because faculty were concerned not with whether the student could get the correct answer, but whether the student knew the procedure needed to obtain the correct answer or not. It is made up of 20 questions chosen from the SAT math that had a direct relationship to principles of microeconomics material. This includes percentage change, ratios, proportions, fractions and decimals, order of operations, place value, the area of a triangle, simple exponents, and the graph of a straight line. Questions were specifically chosen to have very simple calculations so that a calculator should not be necessary. In this way, these questions differ from questions used by many of the papers that make up our literature review, which use complicated numbers (Boyle and Farreras, 2015; Bridgeman, 1992; Bridgeman, Harvey, and Braswell, 1995) and are more similar to the lower-difficulty questions used in Katz, Bennett, and Berger (2000). Students have thirty minutes to complete the assessment. A score of 16 out of 20 implies mastery of basic math skills. Student are informed two weeks before the start of classes of the assessment, and they are given three attempts to pass. Students receive the full 10% towards their grade if they pass on any attempt and receive a 0% if they fail to pass any attempt. The department provides remedial math support for all students that fail the first attempts.

METHODOLOGY

Experiment

A total of 1,354 students took the BMA during the first week of the fall semester of 2013. These 1,354 students were spread across five different professors. The average total score was 13.5/20 implying that on average students answered 68% of the 20 questions asked correctly. To pass students were required to correctly answer 80% of the questions correctly. This led to a failure rate of 65% on the first attempt at the assessment.

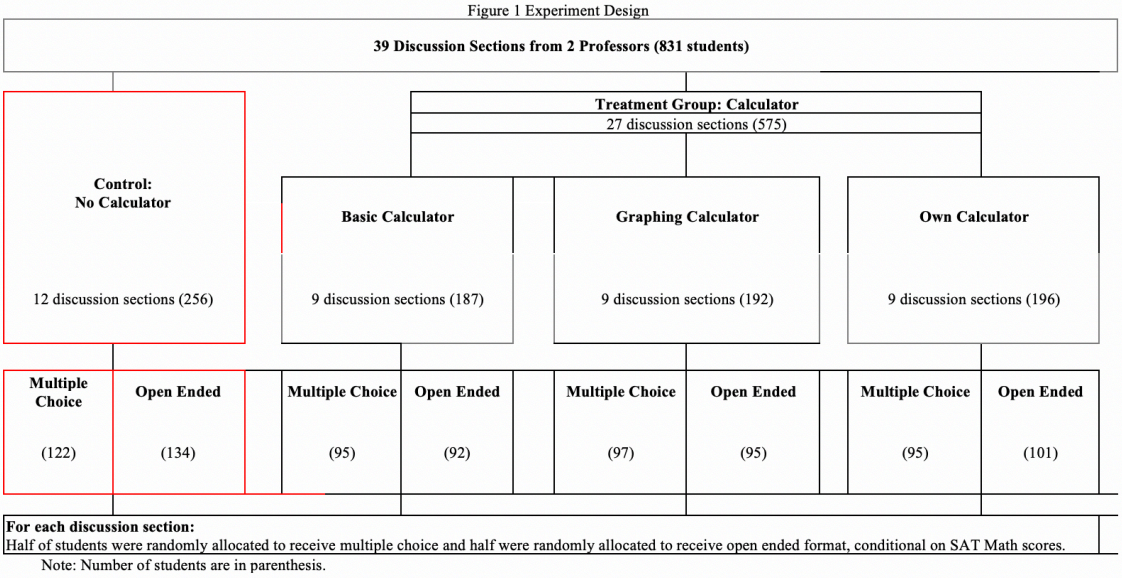
Before administration of the BMA, 902 students from two professors of principles of microeconomics courses were randomly assigned to a treatment or control group, conditional on their SAT math scores. Immediately after the BMA was administered, these selected students were given a second math assessment called the basic math questionnaire (BMQ) containing a sample of 10 questions from the BMA. Students were given 15 minutes to complete the BMQ.

Figure 1 explains the experimental design. Each student that is enrolled in a principles of economics course at this university is assigned to a discussion section which meets once a week with a teaching assistant. Each teaching assistant has three discussion sections. The BMA and BMQ were administered in students' first discussion section meeting in the first week of classes.

Randomization was first done for calculator type at the teaching assistant level. This implies that all of the three discussion sections of the teaching assistant were given a specific treatment or control. Treatment for calculator type was done at this level to ensure that treatment was followed and to ensure that students did not endure unneeded stress by seeing a neighbor using a calculator when he/she was not allowed to use one. Each teaching assistant was also given a helper to ensure that her/his discussion sections would be strictly given the treatment or control. Randomization was conditional on the teaching assistant's students' average SAT math scores to

ensure that treatment and control groups had similar math skill levels. Of the 13 teaching assistants (39 discussion sections), four teaching assistants (including 12 discussion sections with 256 students) were assigned to the control group which did not have the use of a calculator. Three teaching assistants (nine discussion sections) were assigned to each calculator treatment in the form of being given a basic calculator, TI-83 graphing calculator, or allowed to use their own calculator at the beginning of the BMQ.

Figure 1 Experiment Design



Note: Number of students are in parenthesis.

To further test any interactions between calculator use and test format, we randomized within each of the discussion sections (again conditional on students' SAT math scores) such that half of students in each discussion section were given an open ended test framework and the other half of students were given a multiple choice framework. This was done by giving students envelopes with the assigned BMQ inside.

Multiple choice and open-ended questions were chosen from the SAT and were the same except that multiple choice questions had five different possible answers that students could choose from. This is a similar methodology to Boyle and Farreras (2002) and Chan and Kennedy (2002).

In the open-ended framework, students were given a box to provide their numerical answer. In the multiple choice framework, students were asked to circle the correct answer.

The basic calculators used in this assessment were the same as those available to be used by the university's economics faculty. Only addition, subtraction, multiplication, and division can be done with these calculators. For students that were allocated into the treatment where they could use their own calculator, students were asked to report if the calculator was a basic calculator or graphing calculator. It was reported that 40% of students used a basic calculator, 56% used a graphing calculator, and four percent used no calculator. All students were advised prior to the assessment that they should bring a calculator; however, it is likely that many students forgot to bring them on the day of the assessment. No checks were done for possible functions in students graphing calculators brought to the BMQ, whereas the graphing calculators given to students had no functions downloaded onto them.

Of the 902 students that participated in the experiment, 71 students did not consent for their information to be used in the experiment and 34 did not provide additional student characteristics, leaving us with 797 students in our sample. Table A1 demonstrates that the share of students that did not consent does not significantly differ across treatment and control groups. This ensures that there is no bias in the results from students not wanting to participate based on their assignment into a specific treatment or control.

The validity of the experimental design can be seen in Table A2, where statistically significant differences are tested between treatment and control groups. There is no statistically significant difference in the average SAT math score across control and treatment groups. Because not all students took the SAT math, validity was also tested for ACT math scores. There were no statistically significant differences in ACT math scores between treatment and control groups.

Validity of the experiment was also checked when separating students into low-performing (those with an SAT score below the mean) and high-performing. The validity of the experiment continues to hold.

Table 1 provides summary statistics for the 797 students included in the experiment. The first column gives summary statistics for the whole sample and the second column provides summary statistics for only low-performing students who had an SAT math score below the mean score of 662 (358 students). A significant majority of students are freshmen. 40% of students attended private high school, and roughly half of students are female. Female students are slightly more likely to have a lower SAT math score than the mean.

Table 1: Summary Statistics

	Whole Sample	Low Performers
SAT Math	662	604
Freshman	87%	87%
Sophomore	10%	11%
Junior	2%	2%
Senior	.4%	.3%
Private High School	39%	42%
Female	51%	56%
BMA (10 experiment questions)	6.8	5.75
Question 1	90.3%	84.1%
Question 2	97.4%	96.4%
Question 3	48.1%	33.9%
Question 4	45.9%	28.8%
Question 5	72.1%	59.7%
Question 6	61.4%	51.6%
Question 7	47.9%	32.6%
Question 8	80.0%	69%
Question 9	60.5%	45.2%
Question 10	65.1%	57.7%
N	797	358

The average total score on the ten assessment questions relevant for the 797 students included in the experiment was 6.8/10, implying that 68% were answered correctly. Table 1 gives

the percentage of students that correctly answered each of the questions on the assessment Question one asked students to calculate a percentage change and two asked a question related to fractions. Performance was high on these questions. On question one, 90.3% of students answered the question correctly and on question two, 97.4% of students answered the question correctly. Questions three and five asked students to answer algebraic word problems and questions six asked students to find the zero of a function. Students struggled with these questions. Only 48.1% of students correctly answered question three and 72.1% of students correctly answered question five. Questions seven and nine asked students to calculate the area of a triangle. Performance was poor on this section. Question eight asked students to calculate the slope of a line. Question ten covered data analysis, asking students to interpret a pie chart. 65.1% of students correctly answered question ten. Across all questions, low-performing students performed more poorly than the sample as a whole. All questions used in the experiment are available to be viewed in the appendix.

Empirical Strategy

To analyze the impact of each type of calculator (basic, graphing, and own calculator), test format (multiple choice vs. open-ended) and the interaction of test format and calculator use/type on students' ability to answer basic math questions related to material in principles of economics, we use a difference in difference framework as follows:

$$\begin{aligned} Score_{it} = \beta_0 + \beta_1 d_t + \beta_2 MC_{it} + \beta_3 MC_{it} \times d_t + \sum_{j=1}^{3} \beta_{4j} Calculator_{jit} + \beta_{5j} Calculator_{jit} \times d_t \\ + \beta_{6j} Calculator_{jit} \times MC_{it} + \beta_{7j} Calculator_{jit} \times MC_{it} \times d_t \\ + \beta_8 StudentCharacteristics_i + \epsilon_{it} \end{aligned}$$

where i refers to student i, t refers to time period t (where there are two time periods, 0- when students took the BMA and 1- when students took the BMQ), j refers to the type of calculator

assigned ($j = 1$ if the student was provided the basic calculator, $j = 2$ if the student was provided the graphing calculator, and $j = 3$ if the student was allowed to use their own calculator) $Score_{it}$ is the score of student i from the BMA and BMQ (on the assessment - out of 10 points- and on each individual question- 0 if incorrect and 1 if correct), d_t is a dummy variable equal to 1 when t=1 and 0 otherwise, MC_{it} is 1 if the student was randomly allocated to receive the multiple choice framework and 0 if open-ended, and $Calculator_{jit}$ is 1 if the student was randomly allocated to receive the specific type of calculator and 0 otherwise. $StudentCharacteristics_i$ includes the SAT math score, gender, class, and public or private high school attendance of student i. If a student took the ACT math rather than the SAT math, the ACT score was transformed to the equivalent SAT score using the national percentile rankings for the ACT and SAT (ACT, 2013; The College Board, 2013). This is the same method used in Dorans (1999) and Fennell and Foster (2019).

The difference in difference framework is used here because of the nature of the experiment since we can assume that the difference in the change in students' performance between the BMA and BMQ across treatment groups stems only from the treatment. This form also allows us to test for the interaction of the two treatments, taking into account any covariates. The coefficient on the interaction term between calculator use and the multiple choice framework would tell us if there is an additional benefit (or cost) from having both treatments beyond the exclusive benefit of the multiple choice and exclusive calculator benefit.

RESULTS

Results for the impact of calculator use and multiple choice format on students' total scores can be seen in the first column of Table 2 below. We find that introducing multiple choice without the use of a calculator improves a student's score by approximately one point. We also find that

allowing a student to use any type of calculator in an open-ended test framework improves a student's score by one-half of a point. There is no significant difference in the type of calculator used. This implies that a student that received the multiple choice and a calculator had an improvement in their score of one-and-a-half points. This would result in an increase in the average score of all students from 6.8 to 8.3. There is no interactive effect of allowing a student to use both a calculator and multiple choice framework, meaning there is no significant additional benefit (or cost) of both treatments beyond the improvement of the one-and-a-half points. The effect of allowing a student to use multiple choice is significantly greater than the effect of allowing a student to use a calculator.

The regression was also run for each individual question, such that $Score_{it}$ was measured as zero if the student answered the question incorrectly and one if the student answered correctly to analyze the impact of calculator type/use and test format on different types of questions. Probit difference in difference is used to estimate these regressions. Results for all students can be found in columns two through seven in Table 2. Only questions where there was a significant effect from treatment were reported. Questions two, four, five, and nine had no significant impact from either multiple choice framework or calculator use. Questions one, three, six, seven, eight, and 10 all had a significant positive impact from multiple choice framework, and questions one, three, and six had a significant positive impact from calculator use.

Question one asked students to calculate a simple percentage change. Results from Table 2 Column 2 show that giving students a multiple choice framework without the use of a calculator increases their probability of answering the question correctly by 4.4%. Giving students a basic calculator in the open-ended framework increases their probability of answering the question correctly by 3.6% with no effect from any other calculator. This implies that students that had both

the multiple choice framework and basic calculator increased their probability of answering the question correctly by eight percent. Originally, 90% of students answered this question correctly.

Table 2: Treatment Effects for All Students' Total Score and Individual Questions[a][b]

Treatment	Total Score	Question 1	Question 3	Question 6	Question 7	Question 8	Question 10
Multiple Choice	0.929***	0.044**	0.245**	0.129*	0.154*	0.093**	.219***
	(0.283)	(0.019)	(0.076)	(0.079)	(0.094)	(0.047)	(0.056)
Basic Calculator	0.598**	0.036**	0.224***	0.048	0.086	0.020	0.003
	(0.301)	(0.015)	(0.074)	(0.090)	(0.101)	(0.060)	(0.089)
Graphing Calculator	0.590*	0.012	0.191**	0.106	-0.013	0.037	-0.057
	(0.300)	(0.026)	(0.078)	(0.078)	(0.10)	(0.055)	(0.087)
Own Calculator	0.529*	0.002	0.141*	0.209***	-0.05	0.057	0.018
	(0.294)	(0.029)	(0.085)	(0.056)	(0.098)	(0.047)	(0.075)
Multiple Choice x Basic Calculator	-.388	-0.095	-.185	-.033	-0.069	-0.017	0.035
	(0.429)	(0.144)	(0.15)	(0.143)	(0.138)	(0.107)	(0.120)
Multiple Choice x Graphing Calculator	-.093	0.033	0.0166	-.099	-0.048	-0.003	-0.016
	(0.426)	(0.021)	(0.143)	(0.148)	(0.139)	(0.098)	(0.128)
Multiple Choice x Own Calculator	-.319	0.032	-.119	-.074	0.089	-0.066	-0.054
	(0.425)	(.022)	(0.148)	(0.144)	(0.143)	(0.120)	(0.140)
Student Charact-eristics Included	yes	yes	yes	yes	yes	yes	yes
N	797	797	797	797	797	797	797

Questions three and five are both algebraic word problems that require students to set up two equations and solve for two unknown variables. For question five, there is no significant effect from allowing a student to use multiple choice or a calculator. Question five was significantly less

[a] Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

[b] Coefficients in columns two through seven represent marginal effects for the respective probit difference in difference regressions.

difficult than question three, with 72% of students answering question five correctly and only 48% of students answering question three correctly. For question three, giving students multiple choice increases the probability that a student answers correctly by 25%. Giving students a calculator increases the probability that a student answers correctly by 20%, with no significant difference between the type of calculator used. The impact from a calculator and multiple choice are similar. This contrasts the results of Bridgeman, Harvey, and Braswell (1995), Boyle and Farreras (2015), and Schwarz et al. (2002) which found no impact of calculator use on algebraic word problems. There is no significant interactive effect from having multiple choice and using a calculator.

Given that both question three and five were algebraic word problems, the difference in our results suggest an increasing return to calculators and multiple choice framework with increasing difficulty level. This is similar to what is suggested in Katz, Bennett, and Burger (2000), that as the difficulty level increases multiple choice has a larger positive effect. Our results suggest a similar effect for calculators, that as the level of difficulty of the algebraic word problem increases, calculator use has a larger positive effect.

Question six asked students to calculate the zero of a function. Originally, only 61% of students answered this question correctly. Results from Table 2 Column 4 show that giving students multiple choice alone increases the probability that a student answers correctly by 13%. Allowing students to use their own calculator on this question increases the probability that a student answers correctly by 21% with no significant impact from other calculators. This result could be because students had additional applications downloaded onto their own calculator that allowed them to solve the problem. Student-owned calculators were not checked for any additional applications. It could also be that the result stems from students' high level of comfort with their own calculator and are, therefore, better able to use the calculator's capabilities. Therefore,

students with multiple choice and their own calculator would see a 34% increase in the likelihood of answering this question correctly. There is no statistically significant effect of any other calculator, and there is no significant additional effect of using both multiple choice framework and a calculator.

Question seven asked students to calculate the area of a triangle cut from a rectangle. Originally, only 48% of students correctly answered this question. By allowing students to use multiple choice on this question, this increases the probability that a student answers correctly by 15%. According to Table 2 Column 5, there is no significant effect of allowing for the use of a calculator or interactive effect between test framework and calculator use. This result is similar to that of Bridgeman, Harvey, and Braswell (1995) who find that there is no effect of using a calculator when calculating the area of a triangle.

Question nine also asked students to calculate the area of two triangles and calculate the difference in areas. Originally, only 61% of students correctly answered this question. For question nine, there is no significant effect from allowing a student to use multiple choice or a calculator. This may suggest an increasing return to multiple choice framework with increasing difficulty level when calculating areas.

Question eight asked students to calculate the slope of a given line. Originally 80% of students answered this question correctly. Giving students a multiple choice test framework increases the probability that a student answers correctly by nine percent. Giving students the use of a calculator has no significant impact on their probability of answering the question correctly. There is no statistically significant interactive impact from adding a calculator to the multiple choice framework.

Lastly in question 10, students were asked to interpret a pie chart and give the number of people in a given group represented in the pie. Originally only 65% of students answered this question correctly. We find that giving students multiple choice increases the probability that a student answers correctly by 22%. There is no statistically significant impact from having a calculator. There is no statistically significant interactive impact from having any type of calculator with the multiple choice.

In question 10, the most common incorrect answer given on the open ended framework was 50. 50 is the correct percentage of people in the group, but the question asks for the number of people, not percentage of people. The multiple-choice version of the question (taken from the SAT math) did not have the answer 50. Students clearly did not read the question correctly, and then upon looking at the potential options realized their mistake, they re-read the question. Bridgeman (1992) has a long discussion of the negative ramifications of multiple choice use when the most common critical response answer is not included in the multiple choice. In economics, if a percentage is interpreted as a value (for instance maybe an elasticity as a quantity or price), the answer is completely incorrect. In economics, interpretation is key, and giving students multiple choice may not allow for the testing of this interpretation.

The same regression is run for low-performing students with an SAT math score below the mean score of 662. Results can be found in the first column of Table 3 below. The effect of having a calculator and multiple choice on the student's total score becomes larger. Now, utilizing multiple choice questions without the use of a calculator improves a students' score by one-and-a-half points, and allowing a student to use any type of calculator in an open-ended test framework improves a student's score by one point. This implies that a student that received the multiple choice and a calculator had an improvement in their score of two-and-a-half points. This would

result in an increase in the average score of low-performing students from 5.75 to 8.25. Both Bridgeman (1992) and Schwarz et al. (2002) also find that weak students tend to benefit more from multiple choice questions and calculator use.

Table 3: Treatment Effects for Low-Performing Students' Total Score and Individual Questions [cd]

Treatment	Total Score	Question 1	Question 3	Question 6	Question 7	Question 8	Question 10
Multiple Choice	1.448***	0.134**	0.308**	0.216*	0.107	0.132	.325***
	(0.446)	(0.036)	(0.130)	(0.134)	(0.14)	(0.107)	(0.092)
Basic Calculator	0.790*	0.055	0.390***	0.09	0.087	0.096	-0.055
	(0.497)	(0.056)	(0.117)	(0.154)	(0.177)	(0.110)	(0.151)
Graphing Calculator	0.992**	0.07	0.369***	0.212*	-0.101	0.115	-0.113
	(0.475)	(0.045)	(0.116)	(0.127)	(0.126)	(0.103)	(0.154)
Own Calculator	1.031**	0.06	0.269*	0.262**	-0.093	0.064	0.079
	(0.479)	(0.042)	(0.114)	(0.119)	(0.133)	(0.118)	(0.134)
Multiple Choice x Basic Calculator	-.699	-0.39	-.251	-.088	-0.156	-0.208	-.012
	(0.689)	(0.342)	(0.186)	(0.237)	(0.161)	(0.245)	(0.229)
Multiple Choice x Graphing Calculator	-.506	-0.039	0.003	-.33	0.27	0.018	-0.222
	(0.683)	(0.173)	(0.235)	(0.199)	(0.227)	(0.19)	(0.246)
Multiple Choice x Own Calculator	-.800	0.00	-.092	-.301	0.209	0.092	-0.387
	(0.685)	(.000)	(0.229)	(0.216)	(0.235)	(0.157)	(0.233)
Student Charact-eristics Included	yes	yes	yes	yes	yes	yes	yes
N	358	358	358	358	358	358	358

The results from running the same probit difference in difference regressions from Table 2 columns 2 through 7 for only low-performing students are in Table 3 columns 2 through 7. All statistically significant coefficients increase in size thereby bolstering the results from Table 3

[c] Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

[d] Coefficients in columns two through seven represent marginal effects for the respective probit difference in difference regressions.

Column 1 that low performers benefit more from multiple choice and calculator use. The only exceptions are for the coefficients for questions seven and eight in columns 5 and 6 where the coefficients are no longer significant but the relative magnitude remains the same. Question seven asks students to calculate the area of a triangle cut from a rectangle and question eight asks students to calculate the slope of a line. Given the similar magnitude of the coefficients across Tables 2 and 3 for these questions and smaller sample size of low-performing students, the effect of giving students multiple choice may not differ across performance levels for these types of questions. This is a potential area for future research.

CONCLUSION

Research has found that students' performance on a basic math assessment or quiz has a statistically significant positive correlation with performance in principles of economics (Ballard & Johnson, 2004; Fennell & Foster, 2019; Schuhmann et al., 2005; Douglas & Sulock, 1995; Hafer & Hafer, 2002). It has also been found that test format as well as calculator use have a significant impact on college students' ability to correctly answer basic math questions (Frederiksen, 1984; Bridgeman, 1992; Katz, Bennett, and Burger, 2000; Chan and Kennedy, 2006; Boyle and Farreras, 2015; Koop, 1982).

Economics faculty at this university administer a common BMA at the start of the fall semester of principles of economics to assess the math skills of incoming students. The framework is open-ended, and no calculator is allowed. A randomized experiment was run after the BMA was given, where students were randomized into treatment groups based on calculator use and multiple choice format to determine if test format and calculator use have a significant impact on their performance as well as if there is any interactive effect of these two treatments.

We find that calculator use and test format have a significant positive impact on students overall scores. Students that receive multiple choice tests improve their score by one point. Students that received a calculator improved their score by half a point. In combination, students that received both of these improved their score by one-and-a-half points on average. For a 10-point assessment with a cutoff of eight points where the average student scored a 6.8, this implies that the failure rate would decrease from 60% to 39%. When focusing on only low-performing students, those that received a calculator improved their score by one point. Low performers that received the multiple-choice framework improved their score by one-and-a-half points. In combination, low-performing students that received both of these improved their score by two-and-a-half points. For low-performing students, the failure rate would decrease from 83% to 40%.

We find that the impact of each type of calculator and test framework differs depending on the question. We find a positive effect of multiple choice over a range of math problems that are critical for students to know for a principles of economics course. Calculator use has a significant effect on a few problems, specifically algebraic word problems with a high level of difficulty. In this type of problem, there is no evidence that the effect differs by the type of calculator used. We find that the use of different calculators does have an impact. For instance, having one's own calculator when solving for a zero value for a function significantly increases the likelihood of answering a question correctly. This is most likely due to the fact that students know the functionality of their own calculator. These effects are larger for lower-performing students. We find no additional interactive effect of using a multiple-choice test framework and a calculator.

These results suggest that if a university implements a basic math assessment for a principles of economics course, it must understand the potential effects of using multiple choice or allowing a calculator on certain problems. This is especially important for low-performing

students who the assessment is trying to identify and improve the skills of to ensure that economic principles can be adequately understood. Future research could ascertain if students that rely on multiple choice and a calculator to solve basic math problems underperform those that can adequately solve these problems without the use of these tools.

References

ACT, Inc. (n.d.). ACT profile report-national-graduating class 2013. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/Natl-Scores-2013-National2013.pdf

Allgood, S., Walstad, W.B., & Siefgfried, J. J. (2015). Research on Teaching Economics to Undergraduates. *Journal of Economic Literature*, *53*(2), 285-325.

Ballard, C. & Johnson, M. (2004). Basic Math Skills and Performance in an Introductory Economics Class. *Journal of Economic Education*, *35*(1), 3–23.

Becker, W. & Johnston, C. (1999). The Relationship between Multiple Choice and Essay Response Questions in Assessing Economics Understanding. *Economic Record*, *75*, 348–357.

Boyle, R.W. & Farreras, I.G. (2015). The Effect of Calculator Use on College Students' Mathematical Performance. *International Journal of Research in Education and Science, 1*(2), 95-100.

Bridgeman, B. (1992). A Comparison of Quantitative Questions in Open-Ended and Multiple-Choice Formats. *Journal of Educational Measurement, 29*(3), 1992.

Bridgeman, B. & Potenza, M. (1998). Effects of an On-Screen versus Bring-Your-Own Calculator Policy on Performance on the Computerized SAT I: Reasoning Test in Mathematics. Presented at the Annual Meeting of the National Council on Measurement in Education (NCME).

Bridgeman, B., Harvey, A., & Braswell, J. (1995). Effects of Calculator Use on Scores on a Test of Mathematical Reasoning. *Journal of Educational Measurement*, *32*(4), 323–340.

Buckles, S. & Siegfried, J.J. (2006). Using Multiple-Choice Questions to Evaluate In-Depth Learning of Economics. *The Journal of Economic Education. 37*(1), 48-57.

Chan, N. & Kennedy, P.E. (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.

Dorans, N.J. (1999). Correspondence between ACT and SAT I scores. *ETS Research Report Series, 1999*(1), 1-18.

Douglas, S. & Sulock, J. (1995). Estimating educational production functions with correction for drops. *The Journal of Economic Education, 26*(2), 101-112.

Durden, G.C. & Ellis, L.V. (1995). The Effects of Attendance on Student Learning in Principles of Economics. *The American Economic Review, 85*(2), 343-346.

Fennell, M. & Foster, I. (2019). The Role of Basic Math Assessment in Improving Student Performance in Economics. *Journal of Economics and Economic Education Research, 20*(2).

Frederiksen, N. (1984). The Real Test Bias: Influences of Testing on Teaching and Learning. *American Psychologist, 39*(3), 193-202

Gallo, A.A. & Johnson, C.K. (2008). Math skills and everyday problem solving. *Journal of Economic and Finance Education, 7*(1), 7-20.

Hafer, G.H. & Hafer, R.W. (2002). Do Entry-Level Math Skills Predict Success in Principles of Economics? *Journal of Economics and Economic Education Research, 3*(1), 3-12.

Hoag, J. & Benedict, M.E. (2010). What influence does mathematics preparation and performance have on performance in first economics classes? *Journal of Economics and Economics Education Research, 11*(1), 19-42.

Jensen, E.J. & Owen, A.L. (2003). Appealing to good students in introductory economics. *The Journal of Economic Education, 34*(4), 299-325.

Katz, I.R, Bennett, R.E., & Berger, A.E. (2000). Effects of Response Format on Difficulty of SAT-Mathematics Items: It's Not the Strategy. *Journal of Educational Measurement, 37*(1), 39-57.

Koop, J.B. (1982). Calculator Use in the Community College Arithmetic Course. *Journal for Research in Mathematics Education, 13*(1), 50-60.

Kroncke, C. & Smyth, L. (2003). The Effectiveness of Using Electronic Slides to Teach Graphical Economic Concepts. *Journal of Economic and Finance Education, 2*(2), 38-46.

Lagerlof, J. & Seltzer, A. (2009). The effects of remedial mathematics on the learning of economics: Evidence from a natural experiment. *The Journal of Economic Education*, *40*(2), 115–137.

Mallik, G. & Lodewijks, J. (2010). Student performance in a large first year economics subject: Which variables are significant? *Economic Papers: A journal of applied economics and policy*, 29(1), 80–86.

Mallik, G. & Shankar, S. (2016). Does prior knowledge of economics and high level mathematics improve student learning in principles of economics? *Economic Analysis and Policy, 49*(1)*,* 66-73.

Milkman, M., McCoy, J., & Brassfield, D. (1995). Some additional evidence on the effect of university math on student performance in principles of economics. *Journal of Research and Development in Education, 28*(4), 220-229.

Mendez-Carbajo, D., Mixon, F.G., Jr., & Asarta, C.J. (2018). Can Economic Education Improve Entry-Level Math Skills? *Perspectives on Economic Education Research, 11*(1), 19-40.

O'Neil, H. F. Jr. & Brown, R.S. (2009). Differential Effects of Question Formats in Math Assessment on Metacognition and Affect. *Applied Measurement in Education, 11*(4), 331-351.

Owen, A. (2012). 32 Student characteristics, behavior, and performance in economics classes. In G. Hoyt & K. McGoldrick (Eds.), *International handbook on teaching and learning economics* (pp 341-350). Edward Elgar Publishing.

Pinter, M. (2014). Writing to enhance understanding in general mathematics courses. *PRIMUS, 24*(7), 626-636.

Pozo, S. & Stull, C. (2006). Requiring a math skills unit: Results of a randomized experiment. *The American Economic Review*, 437–441.

Rebeck, K. & Asarta, C. (2012). Methods of Assessment in the College Economics Course. In G. Hoyt & K. McGoldrick (Eds.), *International handbook on teaching and learning economics* (pp 177-187). Edward Elgar Publishing.

Scheuneman, J.D., Camara, W.J., Cascallar, A.S., Wendler, C., & Lawrence, I. (2002). Calculator Access, Use, and Type in Relation to Performance on the SAT I: Reasoning Test in Mathematics *Applied Measurement in Education*, *15*(1), 95-112.

Schuhmann, P., McGoldrick, K., & Burrus, R. (2005). Student quantitative literacy: Importance, measurement, and correlation with economic literacy. *The American Economist, 49*(4), 49-64.

Schwarz, R., Rich, C., Areons, E., Podrabsky, T., & Cook, G. (2002). An Analysis of Differential Item Functioning

Based on Calculator Type. *US Department of Education Educational Resources Information Center (ERIC)* Presented at the April 2002 Meeting of NCME.

Smyth, L & Kroncke, C.)., Jr. (2005). Principles of Economics and the Reinforcement of Math Skills. *Shaping the Learning Curve: Essays on Economic Education.* F.G. Mixon, Jr. (ed.). New York, NY: iUniverse, pp26-39.

The College Board. (2013). SAT percentile ranks-2013 college-bound-senior-critical reading, mathematics, and writing percentile ranks. *The College Board.*

Appendix

Table A1: Share of Students that Did not Consent Across Treatment and Control[e]

	Mean in Open Ended No Calculator	*Difference between* Multiple Choice No Calculator *and* Open Ended No Calculator	*Difference between [….] and* Open Ended No Calculator			*Difference between [….] and* Multiple Choice No Calculator		
			Open Ended with basic calculator	Open Ended with graphing calculator	Open Ended with own calculator	Multiple Choice with basic calculator	Multiple Choice with graphing calculator	Multiple Choice with own calculator
	1	2	3	4	5	6	7	8
Non-consent	0.056	0.026	0.033	0.03	0.017	-0.023	-0.034	-0.024
		(0.0307)	(0.033)	(0.033)	(0.031)	(0.034)	(0.033)	(0.034)

Table A2: Validity of Experimental Design[f]

	Mean in Open Ended without calculator	*Difference between* Multiple Choice without calculator *and* Open Ended without calculator	*Difference between [….] and* Open Ended without calculator			*Difference between [….] and* Multiple Choice without calculator		
			Open Ended with basic calculator	Open Ended with graphing calculator	Open Ended with own calculator	Multiple Choice with basic calculator	Multiple Choice with graphing calculator	Multiple Choice with own calculator
	1	2	3	4	5	6	7	8
SAT Math	653	8.78	8.91	1.92	3.18	-7.37	-14.09	-13
		(9.5)	(10.15)	(10.27)	(10.55)	(10.03)	(10.2)	(10.38)
ACT Math	28.52	-0.126	0.182	0.124	0.288	-0.295	-0.211	-0.35
		(0.83)	(0.777)	(0.85)	(0.75)	(0.863)	(0.809)	(0.78)

[e] Standard errors in parenthesis *** p<0.01, ** p<0.05, * p<0.1

[f] Standard errors in parenthesis *** p<0.01, ** p<0.05, * p<0.1

BMQ:

1. If the price of a computer has decreased from $1,000 to $750, by what percent has the price decreased?

A) 10% (B) 15% (C) 25% (D) 33.33% (E) 50%

2. Multiplying by ½ gives the same result as dividing the number by what?

(A) ¼ (B) ½ (C) 2 (D) 3 (E) 4

3. At a school dance Juan paid $2.05 for 1 soda and 2 pretzels. Keisha paid $1.85 for 1 pretzel and 2 sodas. What is the cost of 1 pretzel and 1 soda?

(A) $.55 (B) $1.30 (C) $1.60 (D) $1.95 (E) $2.10

4. If s≠9, and $\frac{s^2-81}{s-9} = t^2$ what does s equal in terms of t?

(A) $t+3$ (B) $t-3$ (C) $\sqrt{t}-3$ (D) t^2+9 (E) t^2-9

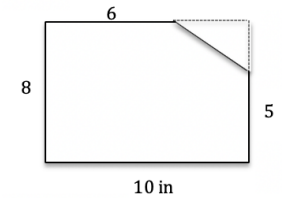
5. Teresita worked 2 weeks in a pharmacy. During the second week she worked 1.5 times the number of hours that she worked during the first week. If she worked a total of 25 hours during the 2-week period, how many hours did Teresita work during the second week?

(A) 5 (B) 7.5 (C) 10 (D) 15 (E) 22.5

6. Let the function f be defined by $f(x) = x^2 - 7x + 10$. If $f(t) = 0$ what is one possible value of t?

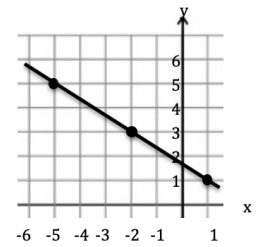
(A) 3 (B) 4 (C) 5 (D) 7 (E) 10

7. (1) A triangular region was cut from a rectangular piece of paper shown below. What is the area in inches of the resulting pentagon?



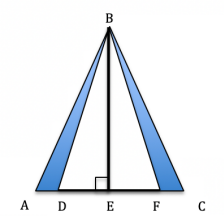
(A) 68 inches (B) 74 inches (C) 86 inches (D) 68 inches squared (E) 74 inches squared

8. (1) The line in the xy-plane below has the equation y=mx+b where m and b are constants. What is the value of m?



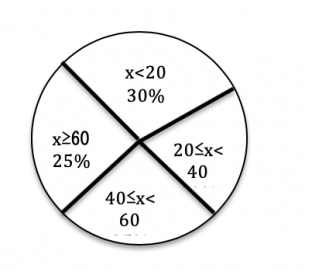
(A) -2/3 (B) -3/2 (C) +1/3 (D) +2/3 (E) +3/2

9. (1) In the figure below, AC=8, BA=BC, BD=BF, BE=10, and AD=1.5. What is the total area of the shaded regions?



(A) 5 (B) 10 (C) 15 (D) 20 (E) 40

10. The chart below shows the results when 1,000 people were asked, "How old are you?" The age they gave is represented by x. How many people said that their age was less than 40?



(A) 200 (B) 300 (C) 450 (D) 500 (E) 550