Clickers and Student Performance: How well do they work for finance students with poorer grades and for students in harder courses?

Scott Anderson, Ryerson University, Toronto, Canada

Allen Goss, Ryerson University, Toronto, Canada

Mike Inglis, Ryerson University, Toronto, Canada

Alan Kaplan, Ryerson University, Toronto, Canada

Laleh Samarbakhsh, Ryerson University, Toronto, Canada

Melissa Toffanin, Ryerson University, Toronto, Canada

Clickers and Student Performance: How well do they work for finance students with poorer grades and for students in harder courses?

Abstract

We study the impact of clickers in two undergraduate finance classes. Consistent with recent literature, we find very little impact on overall performance in clicker vs. non-clicker environments. However, after controlling for course difficulty and student aptitude, we find that clickers have a meaningful positive impact on the performance of weaker (at-risk) students in more challenging quantitative finance courses. Our results suggest that the impact of clickers within the finance context may depend on the type of teaching environment in which they are employed.

Introduction

A considerable body of research evaluates the benefits of clickers (also known as electronic response systems) with respect to student performance. Studies have addressed the issue from the point of view of gender (e.g. King and Joshi, 2008), and small vs. large classes (e.g. Nightingale and Pollack, 2012). Other studies have looked at the effectiveness of clickers based on the subject material (e.g. Tong et. al., 2010; Powell et. al., 2011; Hayter and Rochelle, 2014). However, little or no research has been done to show whether clickers have a smaller, equal, or greater impact on weaker (poorly-performing) students in comparison to students in general, or if the performance impact of clickers is different in courses where students traditionally get lower grades. Arguably, engaging weaker students and/or students in more challenging courses is even more important than other types of engagement. This study attempts to address whether weaker students or students in challenging courses are particularly benefited by clicker use. To address these questions, we measure changes in student performance from the start to the end of a semester in both average and difficult undergraduate finance courses and for both students with poorer grades and other students. Potential benefits of this research include a better understanding of how to work with weaker students so as to help them to succeed and how to develop best practices for more challenging courses.

According to the National Institute on the Education of At-Risk Students (2002) an at-risk learner is any student who may not graduate or who might give up and drop out. The National Institute on the Education of At-Risk Students is part of the U.S. Department of Education. The Institute's purpose is to "carry out a coordinated and comprehensive program of research and development" for the improvement of the education of "at-risk students".[1] We use this mandate as evidence of the emphasis paid to addressing the needs of marginal students. In particular, we recognize that marginal students are likely to be students with poorer grades and students who find challenging courses particular difficult.

[1] Sec. 931(e)(2) of Educational Research, Development, Dissemination and Improvement Act of 1994.

While beyond the scope of this paper, we mention briefly some of the benefits of clicker use to professors. Potential benefits of clicker use in the classroom include taking attendance quickly, administering polls, displaying poll, attendance, or testing results, along with improvements in the ease and speed of grading. In the rest of this paper however, we will focus exclusively on the value of clickers to students, and more specifically to weaker students and students in more challenging courses. Nicholls (1979) and Jackson (1968) both argue that an important marker of successful education for all, but especially for at-risk students, is students' interest in the material. Complementing this position, it is generally well recognized that many students at the college level lack sufficient motivation (McFarlane, 2010) or the academic skills to succeed on their own. The problem can manifest itself on several levels. For example, Beach (2010) says that:

> Many teachers are being faced with difficulty teaching academically diverse students…classrooms may contain students who have no interest in the subject or students who are disrupting class because they are not being stimulated. All this can take valuable minutes away from those students eager to learn. (p.1)

In this paper, we consider the effect of clickers on at-risk students in two ways. First, students can be at-risk because the course that they are taking is difficult. Second, they can be at-risk because they are generally poorly-performing students, whether from lack of motivation, as described above, or for other reasons.

There is a plethora of research that looks at ways to work and/or intervene with weaker students in order to improve their levels of participation, engagement, and performance (e.g. Deslauriers et. al., 2012, and Marx, 2012). The bulk of that research focusses on secondary level education, not post-secondary education. Little of the research considers the impact of clickers on the performance of weaker students in a post-secondary setting, nor on the impact of clickers on the performance of students in difficult courses.

Arguably, clickers and other forms of electronic response systems should be used if doing so improves any of student participation, engagement, or performance in the classroom. Given that students with average or higher grades may already exhibit high levels of participation, engagement, and performance, it seems that interventions that are effective might be most effective for below-average students who could substantially improve their levels of participation, engagement, and performance.

We fill this gap in the literature by asking if and by how much clickers change student performance for weaker students or in classes where the average grade is generally low. Our first research question asks whether the performance gains identified in the literature are monotonic. We conjecture that weaker students can benefit even more than stronger students from the use of clickers. One of our hypotheses is that weaker students will show greater performance improvement than average or stronger student. In part, this may be true because have more potential range to improve. An "A" student can only improve to an "A+", but a "C" student can improve by several grade notches. Second, stronger students are probably already engaged in the classroom and, if so, they cannot improve their engagement significantly since they are already engaged. Conversely, a weak student may be weak for several reasons, one of which may be due to a low level of engagement. Clickers may result in an improvement in that student's level of engagement in the classroom and an improvement in performance, relative to the gains observed for stronger students. Alternatively, the introduction of clickers may result in monotonic performance improvements. This question is an empirical one and we hope that our research results will inform pedagogy accordingly.

Our second general hypothesis is that students in more challenging courses will benefit more from clickers, in comparison to students in easier courses. We will also consider whether weaker students in more challenging courses benefit more than either weaker students in easier courses or average or strong students in more challenging courses. We will more fully define and discuss our hypotheses in Section III of the paper.

Finally, we try to assess the overall effect of clickers based in some way on the criteria set out by Blanc and Martin (1994), who in their examination of the effectiveness of interventions and different strategies for students, provide a number of criteria by which to judge the value of pedagogical techniques. These include:

1. Improvement in academic performance,
2. Persistence of students who elect to participate, and
3. Success generation, which produces confidence, further allowing the student to succeed in other courses/venues. Thus, if clicker use is to be regarded as ultimately beneficial for students, at least one of the first three criteria must be met.

The paper is organized as follows: Section II presents a description of our data collection, Section III sets out our hypotheses, methodology, and results, and Section IV concludes.

Data Collection

For this study, we use performance data obtained from students in two finance courses at Ryerson University in Toronto, Canada. The first course is Managerial Finance (FIN 401) and is required for all undergraduate business students. The second course is Personal Financial Planning (FIN 502). Both courses were delivered in the Fall 2013 term. FIN 401 is a course that has been identified as challenging by students, instructors, and administrators (low average grades, low retention rates), while FIN 502 is a course that has high retention and high pass rates. FIN 401 is also the more analytical and quantitative of the two courses. Thus, for the purpose of this study, we define FIN 401 as the "more difficult" course. Students taking FIN 502 will, for the most part, be students majoring in Finance, and we traditionally observe a relatively high level of average performance in this course relative to the average performance of students in FIN 401, where approximately 75% of the students are not majoring in Finance. Many of these non-Finance students find the required Finance courses to be challenging.

Our data consists of:

1. Pre-test results from students, designed to assess whether a student was poorly-performing or not,
2. Course drop information, designed to determine whether each student stayed in the course until completion, and
3. Final grade results for each student.

We partitioned our sample into classes using clickers and classes using traditional pencil and paper quizzes. For the clicker section, responses to clicker questions accounted for 10% of the final grade. The other section did not use clickers throughout the semester. Instead, four pop quizzes counted for the 10% of the grade. Unlike other studies in the literature, our methodology does not use classes is successive terms or years. Both the treatment and the control group are taught the same material, in the same term, by the same instructor. The only difference is that one section uses clickers while the other uses class does not. This approach attempts to remove the variation that may exist when different exams are used in successive years. In our study, all students wrote a common final exam, regardless of whether clickers were used or not. The same process was undertaken in both FIN 401 and FIN 502.

We report data from 342 of the 513 undergraduate Commerce students who were invited to participate in the study. See Table 1 for descriptive statistics on the number of students in each of the various categories within the study. The remaining students were not included in the study for various reasons, including:

1. They did not give their consent to participate in the study,
2. They did not take the pre-test, and/or
3. They did not stay enrolled in their respective course.

Many students already had clickers from prior courses but those who did not were required to purchase clickers or purchase an application for their smart phone. The cost of the latter was approximately US$15. The cost of data collection from the point of view of the instructor/researcher was relatively low as clicker technology allows instructors to collect student input cheaply. We include this information here because one of the stated aims of Blanc and Martin (1994) is cost effectiveness. Of note, cost-effectiveness increases the likelihood that there was no selection bias in data collection or the self-selection of students into clicker vs. non-clicker sections. In other words, students could not self-select into the clicker vs. non-clicker section in either FIN 401 or FIN 502.

All students, regardless of the course in which they were registered, were given the same pre-test in the second class of the semester. The pre-test was graded and the students' performance was recorded. We use the results of the pre-test to determine the "type" of student. Weaker students are defined as those students who finished in the bottom third of students on the pre-test in each course. For the purposes of this metric, the two FIN 401 sections were grouped together and the two FIN 502 sections were grouped together. For example, of the 216 participating students in the two sections of FIN 401, the 72 students with the lowest pre-test scores were labeled as "Weak", the 72 students with the highest pre-test scores were labeled as "Strong", and the 72 remaining students were labeled as "Average". An identical approach was taken with the two FIN 502 sections.

At the end of the term, each student's final grade in their respective course was recorded. We calculate the difference between a student's final grade and their pre-test score to determine performance, for the purpose of this study. Student performance was then aggregated by course (more difficult or easier), by quality of student (poorer, average, or strong) and by section (clicker or non-clicker), depending upon the research question being addressed.

We also measure performance using a second method. Student retention was tracked in each of the

four classes used in the study. Retention was measured by the percentage of consenting students who stayed in the course through completion. We also chose to consider the retention ratio as a measure of performance. While not a commonly used definition of performance in the literature, course completion data has been used as a measure of student performance in the literature on clickers.[2]

Several protocols are followed in order to best ensure the validity of our results. Both sections of FIN 401 were taught by the same professor in the same semester using the same course outline, the same pedagogy (excluding the use of the clickers), the same pre-test given at the same time, and all students took a common final exam. The two sections of FIN 502 were also taught by the same professor (although a different professor from the one who taught FIN 401) with the same course outline, pedagogy (excluding the use of the clickers), pre-test, and the same final exam. The classes were all of similar size at the start of the study. Students for the clicker and non-clicker sections should have had similar attributes in terms of average ability and range of abilities. Students selected their section prior to knowing that there would be any clickers used – thus, the students were unable to pre-select their sections based on whether clickers were or were not to be used.

Hypotheses, Methodology and Results

In this section of the study, we model and test whether clicker use impacts on student performance. More specifically, we segment the data and then test whether clicker use impacts more greatly on student performance for poorer students (as compared to average or stronger students) and on students in more challenging versus less challenging courses. We propose the following hypotheses:

Hypothesis 1: Weaker students in the clicker sections will improve significantly more than weaker students in the control group (non-clicker sections).

Hypothesis 2: Student performance changes in the clicker vs. non-clicker sections of the less

[2] See, e.g., Kaleta and Joosten (2007).

difficult course will be significantly less than student performance changes in the clicker vs. non-clicker sections of the more difficult course.

The registration and retention statistics for the two courses used in this study are presented in Table 1. There were 270 students registered in FIN 401 at the beginning of the study and 92% completed the course and got a grade. Of these students, 216 agreed to participate in the study, representing 87% participation.[3] The second course, FIN 502 had an initial registration of 243 students and 88% completed the course and got a grade. Participation in this course is lower than in FIN 401, but there is no evidence that this difference in participation impacts our results. Contrary to the evidence reported in Hawkins (2011), there is no significant difference in retention rates between clicker and non-clicker sections of either course in our study.

We report summary statistics in Table 2. The mean final grade in FIN 401 is 66.46%, vs. 73.12% in FIN 502. This difference is significant at the 1% level (at both the mean and the median), and it confirms our characterization of FIN 502 as the easier of the two courses. However, turning our attention to the clicker vs. non-clicker sections, we see a difference between the means and marginal differences at the median for FIN 401 only. Finally, we find no differences in variances between clicker and non-clicker courses, based on Levene's test.[4]

We control for the aptitude of students by segmenting the sample into thirds using pre-test scores. We

[3] Obtaining consent to participate in the study is a requirement of the University. All students were subject to the pedagogy and grading scheme in the section where they registered, regardless of whether they consent. Only students that consent to participate are included in any of the analysis that follows. The research protocol was approved by the Research Ethics Board of the University. Please contact any of the authors of the paper for more details regarding the research protocols.

[4] Levene (1960).

fit a generalized linear model to test whether aptitude (our within-treatments variable), the use of clickers (between treatments), and interactions influence final grades in both courses. Results are reported in Table 3. Panel A of Table 3 reports partial sums of squares for our model. By construction, our within-treatments variable is significant. However, clicker use is only significant at the 5% level in the FIN 401 sample (F=4.04). There is no evidence that clickers influence student success in FIN 502. Based on course content and the nature of the course evaluations, we have defined FIN401 as the more quantitative of the two courses. Given this distinction, our result is similar to the result reported by Lojo (2009) who finds that clicker efficacy is more evident in more quantitative courses. Panel B reports parameter estimates for the generalized linear model. Looking at FIN 401, the weakest students (those with scores in the lowest third of pre-test scores) have final grades 10.5% lower than those in the top third. Students in the middle third have averages 7.1% lower than top students. Similar results are reported for FIN 502. This result is not unexpected, being a function of the experimental design. However, the main effect for non-clicker sections is not significant. Only when interactions are included can we see the differential impact of clickers. The interaction between clicker use and aptitude is isolated to the weakest students, with weak students in clicker sections having 9% higher final grades than those in the non-clicker sections. This effect is limited to the FIN401 class. This effect is significant at the 5% level. There is no statistically significant benefit to the use of clickers for average or stronger students,[5]

Because we find significant interactions, we report simple effects in Table 4. Least squares means are reported for clicker and non-clicker sections in both courses. Consistent with our results in the previous table, when weak students are considered, there is an 8.4% difference between the average scores in clicker and non-clicker sections (63.5% vs. 55.5%), significant at the 1% level. However, this effect is limited to the weaker students. As has been the case with our other findings, there is no evidence of clicker efficacy in FIN 502, regardless of the aptitude of the student.

[5] Top performing students are the omitted group.

In order to verify our results, we winsorized all of the grades at the 5% and 95% level to eliminate any outliers in the data. The results are qualitatively unchanged. Finally, we calculated Cook's Distance, removing all observations with a Cook's D greater than 4/339 (the number of observations in our complete sample). Re-running our model without the influential observations does not qualitatively change our results.

Our evidence supports our first hypothesis that weaker students in the clicker sections will improve significantly more than weaker students in the control group (non-clicker sections). Our evidence also supports our second hypothesis that student performance changes in the clicker vs. non-clicker sections of the less difficult course will be significantly less than student performance changes in the clicker vs. non-clicker sections of the more difficult course.

Conclusion

A substantial amount of research has looked at the effect of clickers on participation, engagement, and student performance. More specifically, many studies have considered the effect of clicker use on exam grades. However, to our knowledge, there exists no research that considers the impact of clickers on the performance of either weaker students or students in difficult courses.

Overall, we find that the use of clickers does significantly improve the performance of weaker students but that the effect is limited to the more difficult course. There is no impact of clicker use either for average or strong students in either course or in any manner for the easier course. Based on our findings, we conclude that clickers can be a valuable tool for students in more difficult courses, and specifically for weaker (potentially at-risk) students.

References

Beach, T.A. (2010). Combining Methodologies in Differentiated Instruction in the Heterogeneous Classroom. *SCMSA Journal*, South Carolina Middle School Association, pp.1-7.

Blanc, R. & Martin, D.C. (1994). Supplemental instruction: increasing student performance and persistence in difficult academic courses. *Journal of the Association of American Medical Colleges*, 69 (6), 452-454.

Deslauriers, L., Harris, S.E., Lane, E., & Wieman, C.E. (2012). Transforming the Lowest-Performing Students: An Intervention that Worked. *Journal of College Science Teaching*, 41(6), 76-84.

Educational Research, Development, Dissemination and Improvement Act of 1994.

Hawkins, L.A. (2011). *Clicker Technology and Course Retention: A Quantitative Study of Maryland Community Colleges* (Doctoral Dissertation), Capella University.

Hayter, J.K., & Rochelle, C.F. (2014). Clickers: Performance and Attitudes in Principles of Microeconomics. *Journal of Economics and Finance Education,* 13(2), 33-44.

Jackson, P.W. (1968). *Life in classrooms*. New York: Holt, Rinehart & Winston.

Kaleta, R., & Joosten, T. (2007). Student Response Systems: A University of Wisconsin System Study of Clickers. *Educause Centre for Applied Research Bulletin*, 2007 (10).

King, D.B. (2010). Redesigning the Pre-exam Review Session. *Journal of College Science Teaching*, 40 (2), 88-96.

King, D.B., & Joshi, S. (2008). Gender Differences in the Use and Effectiveness of Personal Response Devices. *Journal of Science Education and Technology*, (17) (6), 544-552.

Levene, H. (1960). *Contributions to Probability and Statistics.* Redwood City, California: Stanford University Press, 278-292.

Lojo, M. (2009). A comparison of clicker effectiveness for multiple choice and quantitative questions. *California Journal of Operations Management*, 7 (1) 130-135.

Marx, R.W. (2012). Large-scale interventions in science education: The road to utopia?. *Journal of Research in Science Teaching*, 49 (3), 420-427.

McFarlane, D.A. (2010). Teaching Unmotivated and Under-motivated College Students: Problems, Challenges, and Consideration. *College Quarterly*, 13 (3).

Nicholls, J. (1979). Quality and equality in intellectual development. *American Psychologist*, 34, 1071-1084.

Nightingale, J.P., & Pollack, T.A. (2012). The Impact of PRS on Student Performance in Small and Large Classes. *International Journal of Education Research*, 7 (2), 1.

Powell, S., Straub, C., Rodriguez, J., & Van Horn, B. 2011. Using Clickers in Large College Psychology Classes: Academic Achievement and Perceptions. *Journal of the Scholarship of Teaching and Learning*, 11 (4), 1-11.

Robson, J., Robson, M. & Robson, D. (2008). Clickers, Student Engagement, and Performance in an Introductory Economics Course: a Cautionary Tale. *Computers in Higher Education Economics Review,* 20 (1), 4-12.

Sharma, M.D., Khachan, J., Chan, B., & O'Byrne, J. (2005). An Investigation of the Effectiveness of Electronic Response Systems in Large Lecture Classes. *Australasian Journal of Education Techniques*, 21 (2), 137-154.

Tong, L., Premuroso, R.F., & Beed, T. (2010). Does using clickers in the classroom matter to student performance and satisfaction when taking the introductory financial accounting course?. *Issues in Accounting Education*, 26 (4), 701-723.

U.S. Department of Education. (2002). *Office of Education Research and Improvement (OERI)*. National Institute on the Education of At-Risk Students. Washington, DC.

Table 1
Student Retention
This table reports the number of students enrolled in FIN 401 and FIN 502. "Started" indicates the number of students registered at the beginning of the term. "Completed" is the number of students who finished the course and got a final grade. "Consents" indicates the number of students in each section who agreed to participate in the study.

	Started	Completed	% Retention	Consents	% Participation
			FIN 401		
Clicker	135	126	93%	113	90%
Non-Clicker	135	123	91%	103	84%
Total	270	249	92%	216	87%
			FIN 502		
Clicker	132	117	89%	78	67%
Non-Clicker	111	98	88%	48	49%
Total	243	215	88%	126	59%

Table 2
Summary Statistics
This table reports the descriptive statistics for the final grade in two courses, segmented by clicker use. Differences in means are measured by a t-test. Median differences are measured by non-parametric Wilcoxon test. ***, **, * indicate significance between clicker and quiz sections at 1%, 5%, and 10% levels. †††, ††, †, indicate significance at 1%, 5%, and 10% levels between FIN 401 and FIN 502.

	Obs	Min	25%	Mean	Median	75%	Max	Std Dev
FIN 401								
Non-Clicker	103	21.00	55.79	65.33	63.53**	75.33	93.14	12.96
Clicker	113	23.55	59.45	67.50	67.46	73.23	95.95	11.89
Total	216	21.00	57.82	66.46†††	65.54†††	73.75	95.95	12.43
FIN 502								
Non-Clicker	48	51.30	67.82	73.62	73.75	81.82	92.47	9.73
Clicker	78	45.35	65.71	72.84	74.00	81.19	92.16	11.32
Total	126	45.35	65.86	73.12	73.85	81.58	92.47	10.72

Table 3
Generalized Linear Model Results
This table reports the results of a generalized linear model where the dependent variable is final grade and the independent variables are student aptitude, clicker use, and their interactions. Panel A reports Type III sums of squares for the model. Panel B reports parameter estimates. Strong students are the omitted group. ***, **, * indicate significance between clicker and non-clicker sections at 1%, 5%, and 10% levels.

Panel A. Partial Sum of Squares

		FIN 401			FIN 502		
Source	**DF**	**Type III SS**	**Mean Square**	**F**	**Type III SS**	**Mean Square**	**F**
Aptitude	2	5853.63	2926.82	23.32***	1453.34	726.67	7.25***
Clickers	1	507.66	507.66	4.04**	38.92	38.92	0.39
Aptitude*Clickers	2	591.57	295.79	2.36*	251.55	125.77	1.25

Panel B. Parameter Estimates

	FIN 401		FIN 502	
Parameter	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
Intercept	74.308***	2.505	79.414***	1.892
Weak Student	-10.472***	3.193	-13.242***	3.203
Average Student	-7.130**	2.893	-8.982***	2.538
Non-clicker	0.938	3.280	-3.096	3.360
Weak*Non-Clicker	-9.324**	4.439	6.824	5.395
Average*Non-Clicker	-3.404	3.919	6.210	4.281

Table 4
Simple Effects
This table reports the simple effects (interactions) of the generalized linear model. Least square means are reported for each interaction of interest. ***, **, * indicate significant differences between clicker and quiz LS means at 1%, 5%, and 10% levels. Fisher's Least Significant Difference (LSD) at the 5% level are reported for each simple effect.

	FIN 401				FIN 502			
Aptitude Level	**Non-Clicker Mean**	**Clicker Mean**	**Difference**	**5% LSD**	**Non-Clicker Mean**	**Clicker Mean**	**Difference**	**5% LSD**
Weak	55.45	63.84	-8.39***	5.89	69.90	66.17	3.73	8.35
	(2.24)	(1.98)	(2.99)		(3.34)	(2.59)	(4.22)	
Average	64.71	67.18	-2.47	4.23	73.55	70.43	3.11	5.25
	(1.58)	(1.45)	(2.15)		(2.04)	(1.69)	(2.65)	
Strong	75.25	74.31	0.94	6.47	76.32	79.41	-3.10	6.65
	(2.12)	(2.51)	(3.28)		(2.78)	(1.89)	(3.66)	