THE IMPACT OF A PROFICIENCY-BASED ASSESSMENT AND REASSESSMENT OF LEARNING OUTCOMES SYSTEM ON STUDENT ACHIEVEMENT AND ATTITUDES

MICHAEL A. POSNER
Villanova University
michael.posner@villanova.edu

ABSTRACT

This research compares a student-centered, proficiency-based assessment and reassessment of learning outcomes (PARLO) system to traditional assessment in a college-level introductory statistics class. The PARLO class was assessed on learning outcomes using a three-tiered proficiency scale and given the opportunity to resubmit assignments to increase their rating. Students' attitudes towards statistics improved more in the PARLO group, but no differences between groups were found on the CAOS test or on a common final exam. Within the PARLO group, students with a higher resubmission rate scored better on the final exam and those who resubmitted and achieved proficiency performed similarly to those achieving proficiency with the first submission. Assessing proficiency on specific learning outcomes allowed both students and the instructor to better evaluate learning.

Keywords: *Statistics education research; Formative assessment; Assignment resubmission*

1. INTRODUCTION

1.1. BACKGROUND

The current system of assessment in most classes has failed to foster student learning and engagement, particularly in science, technology, engineering, and mathematics (STEM) content areas. Educators' and policymakers' attempts to address low student engagement and performance in STEM through changes in curriculum, instruction, and support programs have had limited impact on student achievement. The traditional structure of assessment is further complicated by the burden on individual classroom practices to yield to the demands of ambitious curricula at the expense of deep student learning and understanding. In an effort to address these assessment challenges, and to improve student learning, educators have responded by experimenting with and adopting a variety of formative assessment strategies designed to identify students' weaknesses followed by a brief period of targeted instruction.

This study builds on current research on formative assessment and examines an alternate structure for assessing performance in an introductory statistics course. This new structure utilizes a proficiency-based assessment program that affords students multiple opportunities to learn, relearn, and be reassessed on a set of learning outcomes with the overriding objective of attaining proficiency. The system was designed to transform the ways students engage with course content and the ways in which students are assessed. If shown to be effective, these classroom changes will contribute to

Statistics Education Research Journal, 10(1), 3-14, http://www.stat.auckland.ac.nz/serj

students' ability and confidence in performing quantitative tasks, eagerness to learn more about statistics, and become equipped to be an active citizen in today's data-driven and evidence-based society. These methods should also be explored to determine whether they generalize to other fields and educational settings.

This study responds to the growing call for substantive quantitative research in educational research. Too often, educational methods are implemented and fostered using only anecdotal evidence (Becker, 2004). In today's society, we use evidence-based medicine and evidence-based technology. In 2007, the American Statistical Association, recognizing the need for more evidence-based education, produced a report based on a series of workshops called *Using Statistics Effectively in Mathematics Education Research* (2007). This report has begun the conversation among mathematics and statistics educators on using research to improve our teaching. This research serves as a model of using such evidence-based research to improve educational methods.

Prior to this study, a pilot study was conducted to refine the assessment system. An overview of these results was previously presented at the 2007 International Association of Statistics Education Satellite Conference on Assessing Student Learning in Statistics and published in the proceedings (Posner, 2007).

1.2. MOTIVATION

This research study is based on the model of the Young Women's Leadership Charter School (YWLCS) of Chicago (http://www.ywlcs.org/), an exemplar of the impact of a paradigm shift of assessment on student learning and engagement. The structural barriers that exist in the current education system are discussed by Farrington and Small (2006) and implemented by the YWLCS. Specifically, they mention the failure of the traditional mechanisms used for assigning final grades. People naively assume that percentages and grades across schools follow a common system and that they are measured objectively. Farrington and Small propose that teachers need to be clear and explicit about the learning outcomes for a course and how students are going to be assessed on these outcomes.

In 2000, the YWLCS opened its doors, under the co-direction of Small, to all girls in the Chicago area. It was created in order to address the chronic patterns of academic underperformance, repeated academic failure, and disengagement from formal schooling among high school students across the United States. The focus was shifting away from the grading paradigm that rewarded quick-learners using averages on tests or percentages of point accumulations to generate letter grades and/or final averages. This system, though efficient for teachers, deprived students of engaging in the learning process and gaining credit working at their own pace.

The assessment system at YWLCS involves proficiency-based evaluation of learning objectives. Teachers define learning outcomes for each class, evaluating students' proficiency (or mastery), and allowing students to learn at their own pace. In each class, 12 to 15 learning outcomes are defined and students produce evidence of learning of these outcomes. The quality of the outcomes and evidence are reviewed by groups of teachers. Students receive scores of *not yet proficient*, *proficient*, or *high performance* on each learning objective and on-demand, web-based access to these performance measures is granted to teachers, students, and parents using a software system designed at YWLCS. Students who wish to improve their proficiency rating are given opportunities to resubmit assignments.

By structuring achievement around learning rather than credit accumulation, high schools potentially reduce the dropout rate and ensure that students graduate with the skills and knowledge requisite for success in college and active citizenship. Course credit and promotion at YWLCS are based on proficiency of learning objectives. Course credit is awarded to students who demonstrate *proficiency* (or *high performance*) in 70% of the course objectives. Students with 70% *proficiency* in all first year outcomes are promoted to sophomore status. Promotion to junior, senior, and graduate are done similarly, with 75%, 80%, and 85% proficiencies required, respectively. Special advising sessions and other support measures are offered to students to assist them in navigating the process.

This school has radically transformed the educational model with the goal of inspiring more students to learn STEM content, persist to graduation and pursue post-secondary STEM-related studies. The YWLCS boasts exceptional outcomes. Data from 2005 show an 8% four-year dropout rate (compared to 33% in Chicago Public Schools (CPS) in 2003 (Allensworth, 2005) and a one-year dropout rate in 2005 of 9.8% (CPS, 2010)), a 77% four-year graduation rate (the highest graduation rate among all non-selective schools in Chicago), and 100% of graduates attending college.

1.3. THE ASSESSMENT SYSTEM

This research study builds on the successful model of the YWLCS by conducting an experiment in the college setting on the effectiveness of these methods. There are three components to this assessment system – defining learning outcomes, grading using a three-tiered proficiency system, and the opportunity to relearn material and resubmit assignments. Each is based on research in the field of education.

Learning outcomes have been discussed in the literature and in teacher training extensively, both in K-12 education (Biggs, 1980), higher education (Allan, 1996; Ewell, 2002), and linking the two of them (Wiseman & Knight, 2003). Learning outcomes are "an explicit description of what a learner should know, understand, and be able to do as a result of learning" (Bingham, 1999). Educationalists assert that learning outcomes help students learn more effectively, make it clear what can be gained from a course, help instructors design their materials and choose appropriate teaching strategies, communicate with others, and design assessment tools (Jenkins & Unwin, n.d.).

Identifying learning objectives or outcomes is an important component to student learning as well as for equality across courses and years. Even those who dispute the need for alarm about grade inflation agree that students should be evaluated based on standards rather than some normal curve reflecting the performance of those who happen, by choice or coincidence, to be sitting around them that semester (Kohn, 2002). Defining learning objectives or outcomes also benefits the assessor in developing authentic assessments and evaluations of student learning.

Once the learning outcomes are defined, the second component of this system is assessing them using proficiency-based scoring, rather than numeric grading. Butler (1988) showed that providing feedback to students increased student performance and interest, when compared to either a group receiving numerical grades or a group receiving both numerical grades and comments. Black, Harrison, Lee, Marshall, and Wiliam (2003) also showed that students who received comments on their work have greater achievement than those given letter grades. These studies show that numeric assessments tend to result in lower performance and interest of students.

The third component of this assessment system involves the opportunity for students to learn the material at their own pace and resubmit assignments to demonstrate learning. This is a form of both formative assessment and standards-based or mastery learning (Bloom, 1981). Formative assessment is where feedback on learning activities is used to modify the method of teaching to meet the needs of the learner (Black & Wiliam, 1998a). Black and Wiliam (1998b) conducted a review of 250 journal articles and book chapters and found that formative assessment raises academic standards in the classroom. Assignment resubmission (Becker, 2006; Karavirta, Korhonen, & Malmi, 2006), standards-based learning (Clymer & Wiliam, 2007) and mastery learning (Kulik, Kulik, & Bangert-Drowns, 1990) have all been shown to be effective in student learning and attitudes, particularly in weaker students.

I have termed this model the proficiency-based assessment and reassessment of learning outcomes (PARLO) system. Based on the work on this study, among others, the National Science Foundation has awarded us a $2.4 million four-year grant to evaluate the PARLO system through a randomized controlled trial of 44 high schools in the Greater Philadelphia area.

2. METHODS

This experiment compares the effects of the PARLO system on attitudes and achievement of undergraduate students in an introductory statistics class for non-majors at a medium-sized, liberal arts university in the Eastern United States. Students enrolled in the class without knowing of the existence of the study. The classes were held back-to-back for 75 minutes each on two afternoons weekly. The material was identical and delivered as similarly as possible. In both classes, learning outcomes were distributed at the beginning of the semester along with the course syllabus. These outcomes included topics like "Summarize quantitative data with stem-and-leaf plots, histograms," "Calculate and interpret linear regression equation," and "Use the z-table to calculate percentiles for standard normal distribution." Both classes were given weekly homeworks, two midterm exams, and a final exam. All assessments were common to both classes, including the final exam. The course also required a semester long project. The Control class was taught using traditional assessment techniques by evaluating each assignment with a numeric score. In the PARLO class, learning outcomes were assessed using a proficiency scale of *Mastery* (M), *Proficient* (P), *Developing* (D), or *Not Submitted* (N) on each course outcome. Students in the PARLO group who did not receive a score of *Mastery* on any learning outcome were allowed to resubmit homework assignments one more time to demonstrate increased proficiency in their learning and improve their grade. Homework grading was done by a graduate assistant who graded both classes.

Approval by the university's Institutional Review Board was obtained. On the first day of class, the assessment system was discussed, and students were invited to participate in the research study by signing the informed consent form. Students were informed that if they elected not to participate, the grading system and course requirements would be the same, but their data would not be included in the research study. Only one student chose not to participate in the study, due to a recent family identity theft issue. She agreed to allow her performance data to be included without additional data gathering outside the confines of the course.

Surveys were collected (electronically) on student information. Data were gathered on race (White vs. Other), gender, class year, major or expected major, and whether they took a statistics class in high school.

High school transcripts, college transcripts, and Math SAT (or ACT) scores were obtained by the university's registrar. These data were summarized as variables on high school math GPA (average score of quantitative classes), previous (including concurrent) college mathematics courses, and information on whether students took the second course in the sequence (including grade). When Math SAT scores were not available, they were approximated from Math ACT scores using a conversion from Schneider and Dorans

(1999). Math aptitude scores for eight students could not be obtained, primarily due to students transferring to the University. Math SAT was analyzed as a numeric score and categorized into one of four groups: high (>670), low (<590), medium (590–670), or missing. Math SAT appeared to be the most related to outcomes of interest and was chosen as the primary control variable for analysis.

Student knowledge was evaluated in two ways. The Comprehensive Assessment of Outcomes in Statistics (CAOS) was given as part of the last homework assignment, and credit was given for completing it. The CAOS test is a forty question assessment tool that measures student understanding of statistical concepts typically covered in a first semester course on statistics and has been validated (Cronbach's alpha of 0.77) on a sample of 10,287 students (delMas, Garfield, Ooms, & Chance, 2006). Students also took a final exam which was common between the PARLO and Control classes. The exams were taken in back-to-back periods and students leaving early were observed to make sure there was no collusion.

Student attitudes were examined using the Survey of Attitudes Toward Statistics (SATS) (Schau, Stevens, Dauphinee, & Del Vecchio, 1995). The survey was given after the first class and again as part of the last homework assignment, with credit given to students for completing it. The SATS is a valid and reliable instrument that includes thirty-six attitudinal questions on seven-point Likert scales, including some reverse-coded items. The SATS subscales - affect ("I like statistics"), cognitive competence ("I understand statistics"), value ("Statistics is important"), difficulty ("Statistics is hard"), interest ("I am interested in statistics"), and effort ("I will work hard" (pre) or "I worked hard" (post)) - were then calculated as the average of relevant items. Ramirez, Emmioglu, & Schau (2010) have shown that the SATS is congruent with the expectancy-value model by Eccles and Wigfield (2002). Eccles and Wigfield state that student expectations and values are linked to achievement, performance, and persistence.

Two variables summarizing student resubmission patterns were created for the PARLO group – *proportion resubmitted* and *delayed proficiency*. *Proportion resubmitted* measures the level of engagement with the new system by the proportion of times a student resubmitted their assignment when the opportunity was available (when they didn't achieve *mastery* the first time). *Delayed proficiency* measures whether a student demonstrating mastery or proficiency does so on the first or the second attempt. It was calculated as the number of times a student received a proficient or better score only on the second attempt divided by the number of times the student received a proficient or better score. Learning outcomes in which students failed to achieve proficiency in their two attempts were excluded from this calculation. (So a value of .20 indicates that 20% of a student's proficiency scores came on the second attempt.)

All analyses were done using MINITAB version 16 for Windows. Summary statistics were calculated. Comparisons of central measures were done using Wilcoxon rank-sum nonparametric tests (for medians), independent sample *t*-tests (for means), and non-paired, dependent sample *t*-tests (for SATS data, as matching could not be done due to a technical glitch). Drop rates were compared using a hypergeometric distribution. Multiple linear regression was used to examine and control for potential confounders. An alpha of 0.05 was considered statistically significant for all tests, except in examining the six subscales of the Survey of Attitudes Toward Statistics. In this case, the Bonferroni-corrected comparison-wise significance level of 0.0083 was used to recognize the six simultaneous tests, which corresponds to an experiment-wise significance level of 0.05.

A non-paired, dependent sample *t*-test was used to analyze the pre- and post-SATS scores. The mean difference ($\bar{d}$) is correctly calculated as in the independent or paired test, where d_i is the difference from pre to post ($d_i = y_i - x_i$), y is the post-score and x is the pre-score. However, the variance becomes

$$Var(d_i) = Var(y_i - x_i) = Var(y_i) + Var(x_i) - 2Cov(y_i, x_i)\ .$$

The variance of y can be estimated from the post-data, and the variance of x can be estimated from the pre-data, but the covariance of x and y needs to be estimated from matched data. Data on paired SATS scores were obtained on 841 students from four introductory statistics classes at another university and the correlations from these data were multiplied by the product of the pre- and post- standard deviations to approximate the covariances of variables for students in this study.

Course evaluations were reviewed for qualitative student feedback on the assessment paradigm.

3. RESULTS

Table 1 (below) presents demographic, Math SAT, and high school statistics course taken distributions overall and by experimental group. None of these variables was significantly different between the PARLO and Control groups. Although all students who dropped were from the Control class, implying that students in the PARLO group were more likely to decide to stick with the course, the magnitude of the difference was not beyond what would be found by random chance alone, possibly due solely to the small sample size.

Table 1. Student Demographics and Comparison by Group

	Total	PARLO	Control	p-value
Sample size	61	30	31	
Dropped	2	0	2	NS[1]
Analytic sample size	59	30	29	
Females	71%	60%	83%	NS[2]
White	83%	77%	90%	NS[2]
HS Stat	29%	33%	24%	NS[2]
Math SAT [mean(sd)]	626 (55)	620 (56)	633 (54)	NS[3]
Low (<590)	24%	27%	21%	NS[2]
Med (590-670)	47%	53%	41%	
High (>670)	17%	13%	21%	
Not Avail	12%	7%	17%	

[1] hypergeometric test, [2] chi-squared test of independence, [3] two-sample *t*-test

The primary outcome of interest was score on the CAOS test, with the score on a common final exam being a secondary outcome of interest. The PARLO and Control groups did not differ significantly in the performance on the CAOS test and the final exam overall. This comparison was done of the means (using two-sample *t*-test, $p > 0.10$) and the medians (Wilcoxon rank-sum test, $p > 0.10$). Results were not different by Math SAT ($p > 0.10$). However, males in the PARLO group performed 16 points worse on the common final exam than those in the Control group ($p < 0.05$). A Wilcoxon rank-sum test for medians of males was also compared and found to be statistically significant (100 vs. 86, $p < 0.05$). However, the fact that there were only five males in the Control group calls inference made from this small group into question. In addition, the difference between PARLO males and Control males was not significant when examining the CAOS test.

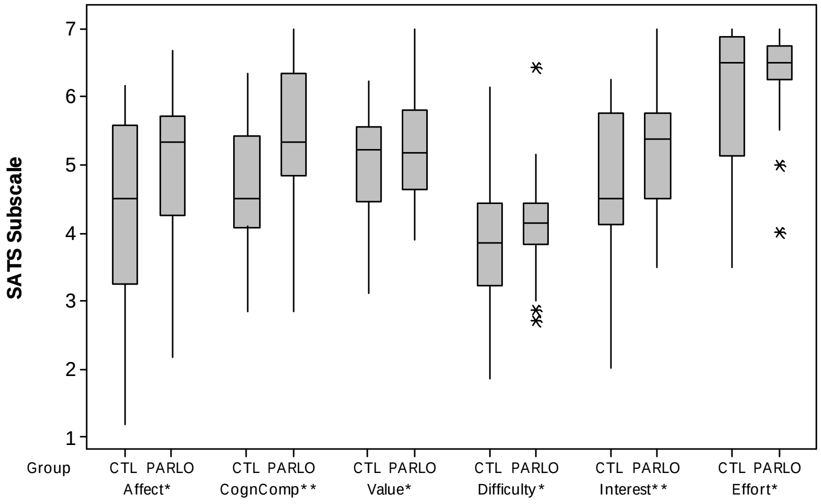
Table 2 (below) presents results from simple and multiple regressions on the CAOS test and final exam by experimental group. Multiple regressions on CAOS were examined controlling for gender and Math SAT. Experimental group was not statistically significantly related to CAOS when controlling for gender and Math SAT. However, the gender-PARLO interaction effect was significant ($p < 0.01$) when including the main effects as well as when additionally including Math SAT. As seen in Table 1 (above), males in the Control group performed very well in the course, both on the CAOS test and on the final exam.

Table 2. Regression results for CAOS and final exam

Dependent Variable	CAOS			Final Exam		
Model	Bivariate	Main Effects	Interaction	Bivariate	Main Effects	Interaction
Intercept	21.6 (0.7)	4.3 (6.1)	3.9 (6.0)	82.6 (2.4)	-7.5 (18.1)	-10.2 (16.2)
PARLO (vs. control)	0.3 (0.1)	0.7 (1.0)	1.5 (1.2)	-0.1 (3.4)	0.2 (3.1)	6.0 (3.2)
Gender (Male)		-0.8 (1.2)	1.2 (1.8)		-1.3 (3.5)	12.3 (4.9)*
Math SAT		0.03 (0.01)*	0.03 (0.01)*		0.14 (0.03)*	0.14 (0.03)*
Gender-PARLO			-3.2 (2.3)			-22.0 (6.1)*
R-squared	0.00	0.15	0.18	0.00	0.36	0.50

Numbers are parameter estimates (standard errors), * indicates p-value < 0.05

Student attitudes were examined using the Survey of Attitudes Toward Statistics (SATS). Figure 1 (below) shows the subscale results from the SATS. All six subscales showed significantly better attitudes among the PARLO group with cognitive competence ("I understand statistics") and interest ("I am interested in statistics") being strongly significant ($p < 0.0001$). This relationship was true even after adjusting the significance level for multiple comparisons (comparison-wise significance level of 0.0083).



*p-value < 0.0083, **p-value < 0.0001

Figure 1. Boxplots of student attitudes (SATS)

CAOS scores were higher among students who chose to resubmit assignments when given the opportunity to do so (see Figure 2). This resulting regression was

Predicted CAOS = 17.7 + 6.2 *proportion resubmitted*, implying that the estimated average score for a student who never resubmitted was 17.7 and a student who resubmitted all their assignments would be 23.9. This relationship was significant (*p* = 0.014) when examined on the CAOS test and strongly statistically significant (*p* < 0.001) for final exam score. It is possible that the proportion resubmitted could be a surrogate marker for diligent students or student aptitude. To examine this, the multiple regression of CAOS on proportion resubmitted controlling for Math SAT was examined. After controlling for Math SAT, this relationship was no longer statistically significant for neither the CAOS nor the common final exam (*p* > 0.1).

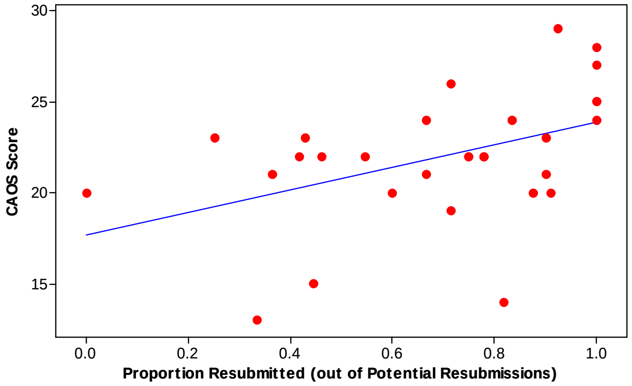


Figure 2. CAOS score vs. proportion resubmitted

Figure 3 (below) shows that students who frequently only achieved proficiency in learning outcomes on their second attempt (high *delayed proficiency*) scored as well on the CAOS as those who more frequently achieved proficiency on their first attempt. As can be seen in Figure 3, there is no association (*p* = 0.73). This lack of association persisted when controlling for Math SAT and gender (*p* = 0.84) A similar result (*p* = 0.90 unadjusted, *p* = 0.84 adjusted for gender and Math SAT) was observed using the final exam score as the outcome measure. Of note, the only student in the PARLO class to receive a perfect score on the final exam was one who required a second submission to achieve proficiency one third of the time.

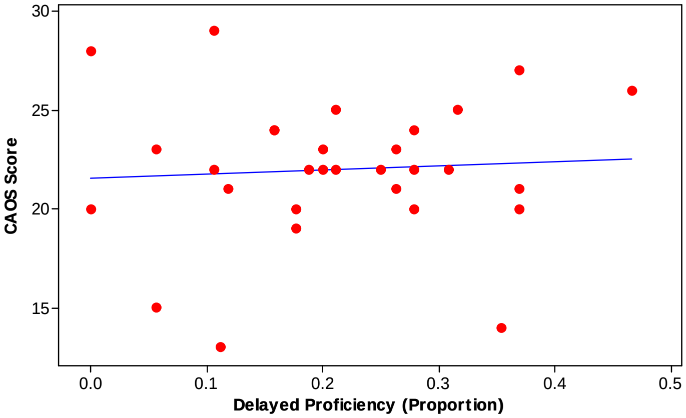


Figure 3. CAOS score vs. delayed proficiency

Student feedback on course evaluations was reviewed for comments relating to the assessment paradigm. Most students liked the grading system. Comments included “You know what to expect each class,” “I like [his] method of grading,” “I found the grading system to be well rounded and fair,” and “I really like the way he graded. It was a very fair way and allowed students more time to review the work, if necessary.” Another student stated that “The ability to resubmit work was an outstanding system.” A few students said they did not like the grading system, with the primary explanation being since “I had no idea where I stood on my grade.”

4. DISCUSSION

The current assessment paradigm used by high school, colleges, and universities is designed more for teacher efficiency rather than centered on paradigms that foster student learning. In order to promote authentic student learning and accurate assessment of that learning, changes in the way we assess students are needed.

There are three major findings from this study. First, students had more positive attitudes towards statistics being in the PARLO class versus the Control class. This has been shown by Eccles and Wigfield (2002) to lead to both achievement and persistence and will likely help to remedy the poor image of the statistical profession as these students move into the workforce. Second, students who achieved proficiency in learning outcomes only on a second attempt showed equivalent performance on the final exam to those who demonstrated proficiency on their first attempt. These students are not ‘second-class’ learners and have demonstrated that they have the ability to perform equally well when given the opportunity to learn the material at their own pace (a second time). Third, students who engage in the process of resubmission of assignments perform better on the final exam (although this may just be a surrogate for math aptitude, because this relationship wasn’t significant when controlling for Math SAT).

No differences between experimental groups were observed on the CAOS test or on the common final exam. In fact, a negative effect was seen for males. This, contrary to results seen at the YWLCS, may have resulted from a number of factors. This study was done in a single class during one semester, whereas the YWLCS uses this paradigm for all its classes over a four year period. Limited feedback was provided to the students in this study, whereas the YWLCS has created an infrastructure to provide individualized feedback to their students. In addition, students might have been reticent being involved in a research study, likely for the first time in their lives. Men in the Control group performed better than men in the PARLO group. It is unclear whether there is a reason for this or whether it is an anomaly.

Defining learning outcomes is an important component of this assessment scheme. Identifying these outcomes helped student learning, instruction, and communication of course content to others, consistent with those stated in the second paragraph of Section 1.3. Evaluation based on learning outcomes better enabled both students and the instructor to assess learning and design good instruments. Having these learning outcomes helped verify that each question on assignments directly related to the topics covered and the skills that students were taught, rather than simply ‘fun’ or ‘interesting’ problems. It also helped facilitate communication between students and the instructor. Rather than coming to office hours frustratingly stating “I don’t understand what’s going on in the class,” students would come saying “I need help with Learning Outcome 5.” This helped provide a more focused and productive dialogue. Rather than complaining about their grade, they asked for help in demonstrating proficiency on a second attempt.

There were flaws in the implementation of this study. Posting proficiency ratings of learning outcomes online was often slow, not allowing students to take full advantage of knowing what they have learned and deciding whether to resubmit. Assignment grading by a graduate assistant may have introduced her biases to the experiment. These factors likely contributed to biasing towards the null, reducing the impact of the intervention.

There are a number of improvements recommended for future research. First, refining the learning objectives to include how they will be measured and providing rubrics of grading to students. Second, a longer-term implementation of this system may allow students to learn the system and succeed. Students at the YWLCS often spend their first year getting used to the new assessment system. Third, the opportunity to resubmit assignments more than one time (as many online homework systems allow) would further allow students to master the material. Fourth, the YWLCS provided additional instruction and interaction after assignments to foster formative learning whereas minimal interactions with the teacher in this study were demanded of the students.

In conclusion, this study begins to demonstrate the usefulness of the PARLO system. It strengthens the call for more research on the impact of this system and its three components—defining learning outcomes, proficiency-based assessment, and the opportunity to resubmit assignments—as well as other changes in instruction and assessment that can improve student learning. Testing new assessment and learning strategies is important. As statisticians and statistics educators, we should generate and use evidence-based teaching methods that both inform and transform our teaching, as well as that of our colleagues. While constantly striving to improve our teaching is often hard and time-consuming, if our goal is to have students learn the material, it is a necessary step in improving our education system.

ACKNOWLEDGEMENTS

Support for this research was provided by the Math Science Partnership of Greater Philadelphia. The author would also like to thank the Young Women's Leadership Charter School and Amy Froelich for their assistance in the research. Neha Chawla and Michael Miller assisted in data processing.

REFERENCES

Allan, J. (1996). Learning outcomes in higher education. *Studies in Higher Education, 21*(1), 93–108.

Allensworth, E. (2005). Graduation and dropout trends in Chicago: A look at cohorts of students from 1991 through 2004. Chicago: Consortium on Chicago School Research.

American Statistical Association. (2007). *Using statistics effectively in mathematics education research (SMER Report).* Alexandria: Author. [Online: http://www.amstat.org/education/pdfs/UsingStatisticsEffectivelyinMathEdResearch.pdf]

Becker, K. (2006, May). *Death to deadlines: A 21st century look at the use of deadlines and late penalties in programming assignments.* Paper presented at WCCCE 2006 - The Western Canadian Conference on Computing Education, Calgary, Alberta. [Online: http://dspace.ucalgary.ca/bitstream/1880/46725/1/Death-to-Deadlines.pdf]

Becker, W. E. (2004). Quantitative research on teaching methods in tertiary education. In W. E. Becker and M. L. Andrews (Eds.), *The scholarship of teaching and learning in higher education: Contributions of research universities* (pp. 265-310). Bloomington, IN: Indiana University Press.

Biggs, J. B. (1980). The relationship between developmental level and the quality of school learning. In S. Modgil & C. Modgil (Eds.), *Toward a theory of psychological development within a Piagetian framework*. Windsor, England: National Foundation for Educational Research Publishing.

Bingham, J. (1999). *Guide to developing learning outcomes*. The Learning and Teaching Institute Sheffield Hallam University, UK: Sheffield Hallam University.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, B. (2003). *Assessment for learning: Putting it into practice.* Berkshire, England: Open University Press.

Black, P., & Wiliam, D. (1998a). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139-148.

Black, P., & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7–74.

Bloom, B. S. (1981). *All our children learning: A primer for parents, teachers, and other educators*. New York: McGraw-Hill Paperbacks.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*(1), 1–14.

Chicago Public Schools, Office of Performance. (2010). *One Year Grade 9-12 Dropout Rates, 1999-2010*.
[Online: http://research.cps.k12.il.us/cps/accountweb/Reports/citywide.html]

Clymer, J., & Wiliam, D. (2007). Improving the way we grade science. *Educational Leadership*, *64*(4), 36-42.

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2006). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, *6*(2), 28-53.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ6%282%29_delMas.pdf]

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109-132.

Ewell, P. T. (2002). An emerging scholarship: A brief history of assessment. In T. W. Banta (Ed.), *Building a scholarship of assessment* (pp. 3–25). San Francisco: Jossey-Bass Publishers.

Farrington, C. H., & Small, M. H. (2006, April). *Removing structural barriers to academic achievement in high schools: An innovative model.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco

Jenkins, A., & Unwin, D. (n. d.). How to write learning outcomes. *National Center for Geographic Information and Analysis Geographical Information Science Core Curriculum Learning Outcomes.*
[Online: www.mf-foca.com/projekti/euser/mostar/how_to%20write_learning_outcomes.doc]

Karavirta, V., Korhonen, A., & Malmi, L. (2006). On the use of resubmissions in automatic assessment systems. *Computer Science Education*, *16*(3), 229–240.

Kohn, A. (2002). The dangerous myth of grade inflation. *Chronicle of Higher Education*, Nov. 8, B7.

Kulik, C. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning: A meta-analysis. *Review of Educational Research*, *60*(2), 265-299.

Posner, M. A. (2007). Evaluating pedagogical techniques in introductory statistics: Proficiency grading and assignment resubmission. In B. Phillips and L. Weldon (Eds.), *Proceedings of the IASE/ISI Satellite Conference on Assessing Student Learning in Statistics*. Voorburg, The Netherlands: International Association for Statistical Education, International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/sat07/Posner.pdf]

Ramirez, C., Emmioglu, E., & Schau, C. (2010, August). *Understanding students' attitudes toward statistics: New perspectives using expectancy-value theory of motivation and the Survey of Attitudes Toward Statistics*. Paper presented Joint Statistical Meetings, Vancouver, BC, Canada.

Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement, 55(5),* 868–875.

Schneider, D., & Dorans, N. (1999). Concordance between SAT I and ACT Scores for individual students [Electronic Version]. *Research Notes,* RN-07, 1-8

Wiseman, D. L., & Knight, S. L. (2003). *Linking school-university collaboration and K-12 student outcomes*. New York: American Association of Colleges for Teacher Education.

MICHAEL A. POSNER
Department of Mathematical Sciences
Villanova University
800 Lancaster Ave.
Villanova, PA 19085