Evaluation of Impact of Calculus Center on Student Achievement

Cameron Byerley
Colorado State University

Travis Campbell
Colorado State University

Brian Rickard
University of Arkansas

Many universities are spending resources to establish math tutoring centers. Sharing information about the effectiveness of such centers is crucial to determine how to allocate resources. We illustrate methods of evaluating tutoring centers. We investigate the question, "what is the association between students' attendance at the Colorado State University Calculus Center and their grade in Calculus II?" We found a statistically significant positive correlation between students' tutoring center participation and their grades.

Key words: Resource Center, Calculus, Student Support, Tutoring, Evaluation

The Characteristics of Successful Programs in College Calculus study (CSPCC) recommended that universities have "proactive student support services" and found that tutoring centers foster "student academic and social integration" (Bressoud, Mesa & Rasmussen, 2015, p.viii) Ninety-seven percent of the 118 US institutions that responded to the CSPCC survey question about tutoring centers had a tutoring center (Bressoud, Mesa & Rasmussen, 2015, p. 70). Tutoring centers in both the UK and the USA are asked to evaluate their success to secure and maintain funding (Personal Communication, Mills, 2017; Matthews et. al. 2012). In addition to evaluation for funding, tutoring centers should be evaluated to determine the "optimal strategies for delivery of support" (Kyle, 2010, p. 104). Tutor training, education level of tutors, format of tutoring, use of technology, location of center and more varies from center to center (Bressoud, Mesa & Rasmussen, 2015; Perin, 2004, p. 563-564).

Conceptual Framework: What counts as success?

It is difficult to measure if tutoring centers achieve their goals using data that is commonly collected. One goal of tutoring centers is that students learn "mathematics worth knowing" (Thompson, 2008, p. 46). We want them to understand calculus as a sensible tool to understand the rate of change and accumulation of real-world quantities (Thompson, Byerley & Hatfield, 2013). However, good scores on calculus tests do not imply students are learning mathematics worth knowing. The CSPCC study collected Calculus 1 final exams from 253 US universities. They found "the exams generally require low levels of cognitive demand, seldom contain problems stated in a real-world context, rarely elicit explanation, and do not require students to demonstrate or apply their understanding of the course's central ideas" (Tallman, Carlson, Bressoud & Pearson, 2016, p. 105).

Another goal of tutoring centers is to help students complete STEM degrees. Centers help students become socially and academically integrated into the university, which helps retain first year students (Bressoud, Mesa & Rasmussen, p. 82; Solomon, Croft & Lawson, 2010; Tinto, 1997). We recognize women are more likely to switch out of a STEM degree even if they are equally qualified as men (Ellis, Fosdick & Rasmussen, 2016).

These goals are important, yet hard to directly measure given data commonly collected. Many centers report the difficulties of both running a tutoring center and gathering and analyzing quality data (Matthews, et. al, 2012). Despite the acknowledged limitations, we define success as a positive correlation between a student's attendance at the Calculus Center (CC) and the student's score in Calculus II after controlling for other variables impacting success. Future studies could consider students' scores on validated assessments on calculus concepts and students' persistence to graduation with a STEM degree.

Literature Review: Evaluating Tutoring Centers

Matthews, et. al (2012) wrote the most complete literature review of evaluation of tutoring centers located in the UK, Ireland, and Australia. We discuss a subset of the studies reviewed, plus additional studies from the US.

Some studies found positive statistical relationships between student success in courses and tutoring center attendance (Dowling & Nolan, 2006; Cuthbert & MacGillivray, 2007; Mac an Bhaird, Morgan & O'Shea, 2009). All of these studies suffer from the difficult to avoid self-selection bias. Students who are more likely to use tutoring center are more likely to share other characteristics that impact grades such as motivation. Some studies used qualitative data to evaluate the effectiveness of centers. For example, Carroll and Gill (2012) qualitatively evaluated a tutoring center using student evaluations.

Not all studies found a positive relationship between tutoring center attendance and grades. For example, Walker and Dancy (2007) found that students who attended a physics tutoring center had 20 percent lower mean exam scores than those who never attended (p. 138). They hypothesized that students who struggled self-selected to use the tutoring center.

Cooper (2010) found that a drop-in multi-subject tutoring center helped increase students' GPAs and persistence in college. Students who came to the tutoring center at least 10 times had on average 0.2 higher GPA and were 10% more likely to persist in college. However, Cooper (2010) did not find a relationship between students' tutoring center attendance and performance in particular courses.

Evaluating the Impact of Tutoring

There have been hundreds of articles about the impact of one-on-one tutoring. Topping (1996) reviewed the literature about peer-tutoring for undergraduate students. Although many studies "suffered from problems of self-selection to groups" (p. 335), Topping found evidence that having advanced undergraduates tutor newer undergraduates improved tutee's grades and was cost efficient (p. 338). Leung (2015) conducted a meta-analysis of studies on peer tutoring in all subjects at the K-16 level. Leung computed a weighted mean effect size, for studies on tutoring, finding a significant weighted mean effect size, d=.43, p<.001, for undergraduate tutoring. This was found to be a larger effect than tutoring at kindergarten and elementary levels but smaller than that for secondary education. Leung found significant effect sizes at all academic ability levels and all school levels, but the meta-analysis does not address the differences in going to a tutoring center versus having a one-on-one tutor.

Colver and Fry (2015) noted "a vast majority of research that is available relies exclusively on correlational, qualitative, or other similarly limiting methodologies that make it difficult to glean insight into the causal impact that tutoring might have on student success" (p. 16). Annis (1983) randomly assigned students to read articles under control, tutor, or tutee conditions. Students who tutored others had significantly greater learning gains than those who were tutored. Lidren and Meier (1990) randomly assigned psychology students to receive frequent, minimal, or no tutoring. They found a statistically significant positive relationship between tutoring and success on class exams. Arco-Tirado, Fernández-Martín, and Fernández-Balboa (2011) randomly assigned undergraduates to receive tutoring on study skills and found that there was no statistically significant relationship between tutoring and success.

Causation versus Correlation

Administrators want to know if tutoring centers or some other intervention is a better use of funds. Tutoring centers would like to show the center *caused* student success. Demonstrating causal relationships requires random assignment to the treatment condition and students can not be randomly assigned to use or not use a tutoring center. A correlation between tutoring center attendance and course grades does not imply a causal relationship because students self-selected to use the tutoring center. It is possible that the weaker students are more likely to self-select to tutoring and that we could expect tutored students to have lower grades (Munley, Garvey & McConnell, 2010;Walker & Dancy, 2012). On the other hand, we could argue that more motivated students are more likely to use tutoring and are also more likely to engage in many other behaviors that will increase their grades. We are not the first to note that many studies of tutoring should include control variables "to rule out the possibility that students with better skills (higher GPA) are more inclined to seek help than those with poorer skills (lower GPA (Perin, 2004, p. 580).

Most of the studies Matthews, et. al. reviewed did not provide evidence that students' improvements in grades were caused by the tutoring center because the studies did not control for self-selection bias. For example, Pell and Croft (2008) used tables to compare the percentage of students who earn various grades and tutor center attendance. They did not control for other variables. MacGillivray and Croft (2011) advocated for tutoring centers to use more rigorous methods to evaluate tutoring center success. They wrote "the essential concept is to compare performance relative to a base measure for those who used [the tutoring center] with the same relative performance for those who did not" (p. 15). They suggested use of students' prior GPA, results on a first assessment, and diagnostic test data as possible baseline measures. They noted two studies that used diagnostic testing as a baseline measure (Dowling & Nolan, 2006; Bamforth, Robinson, Croft & Crawford, 2007). Mac an Bhaird, Morgan, & O'Shea (2009) used students' performance in past school-level examinations as a baseline. Although MacGillivray and Croft (2011) noted that general linear models are useful for analyzing the relationship between many variables and student performance, they only noted one study of tutoring centers (MacGillivray & Croft, 2003) that used general linear models.

Munley, Garvey and McConnell (2010) used the student's high school rank, SAT math score, current college GPA, number of credits the student is enrolled for, freshman or sophomore status, gender, race, participation in Greek life, student attendance of recitation session led by graduate teaching assistants, and course instructor as control variables. They found that students who were tutored did not have statistically significantly different grades. MacGillivray and Croft (2011) also suggested similar control variables and also suggested using a diagnostic test.

General linear models are considered useful in evaluating the impact of education interventions in general. As detailed by Theobald and Freeman (2013), the most commonly used methods to analyze learning gains pre-post test data in undergraduate STEM education -- raw change scores, normalized gain scores, normalized change scores and effect sizes -- fail to control for observable student characteristics; hence, researchers should instead use linear regression to control for observable factors.

Statistical Methods

Colorado State University established a Calculus Center (CC) in August 2016. The tutoring is provided by faculty who teach calculus, graduate teaching assistants, and undergraduate learning assistants who also attend the course that they tutor. The data was collected from four large sections of Calculus II taught by three different instructors.

We will model the average relationship between performance in Calculus 2 and the number of visits to the CC by estimating a generalized linear model (GLM) of the binomial family with a logistic link function. The dependent variable is each student's total score minus attendance, midterm 1 and graph extra credit scores and is used as a measurement of performance. Performance will be modeled as a binary grouped variable: the sum of 636 independent homogeneous Bernoulli trails, implying that Performance has a binomial distribution with parameter 636 (Gujarati & Porter, 2009 p. 557; McCullagh & Nelder, 1989 p. 102). This estimation technique models the sigmodal, non-linear relationship implicit to bounded endogenous variables that is neglected by ordinary least squares estimation.

The parameter of primary interest to this paper is the number of visits to the CC. The initial regression will have six required variables that control for student motivation and mathematical ability and 13 additional test parameters. The required variables include: three diagnostic math questions, midterm 1, attendance, high school GPA, and an indicator of low previous performance (denoted LLP) taking value of one if student reports a C or lower in previous calculus class. The test variables are number of visits to CC, honors section indicator, the number of times the student took Calculus 1, the number of times the student took Calculus 2, the student's total credit hours, honors status indicator, first generation indicator, minority indicator, masters or second bachelors indicator, international student indicator, age, and male indicator. The final models are obtained by running all possible subsets of the test parameters and selecting the model with the least exogenous variables within 2 of the minimum corrected Akaike information criterion (AICc): a change in the AICc that is less than two is negligible (Burnham & Anderson, 2002; Cavanaugh, 2009). AIC is a goodness of fit measure which eliminates the subjective judgment in hypothesis testing (Akaike, 1974). AICc is AIC with a larger penalty for additional parameters. Using the minimum AICc in lieu of p-values is done for predictive accuracy as it minimizes the distance between the true model and candidate model (Burnham & Anderson, 2002).

In estimating student performance, ordinary least squares regression (OLS) is commonly used. However, performance scores are bounded and OLS estimates are not. In general, using OLS with a proportional dependent variable that is bounded between 0 and 1 is only valid if most observations are within 0.3 to 0.7. Approximately 20% of our observations satisfy this criterion. Hence, OLS estimates may entail non-normal and heteroscedastic residuals, unbounded predictions and a reduction in explained variability in the dependent variable (Gujarati & Porter, 2009). Therefore, we use logistic regression with a non-binary response variable. While logistic regression is most commonly used with a binary response variable, it can also be used with a bounded proportion that falls between 0 and 1 as suggested in Papke and Wooldridge (1996). One disadvantage to logistic regression is the large sample requirement relative to OLS. As a general rule, logistic regression requires at least 30 observations per predictor variable. However; some argue that there should be at least 50 observations per predictor variable (Burns & Burns, 2009).

To help address the potential problem of heteroscedasticity (Gujarati & Porter, 2009), we will report two models. The first will use all observations and the second will omit all observations for which $Cooks\ D_i > 3 * Mean(Cook's\ D)$ (Cook, 1977). Cook's Distance (Cook's D) is a measure of each observation's leverage and residual values. It is used to identify influential outliers in a predictor set. By removing influential observations with the Cook's D criteria, we investigate if the estimates are robust to outliers that may be caused by our inability to properly control for previous mathematical ability and motivation.

Students with high Cook's D values correspond primarily to three groups: low performance students, students who checked into CC more than 60 times, and high preforming students without intervention. Points with high Cook's D values should be examined for validity (Stevens, 1984). Some students came to the CC every week between their classes to work on homework for other classes. Some high-performing students never attended the CC because they did not need tutoring. Finally, some students did not use center and showed no signs of effort to pass the class. We are most interested in students who were attempting to pass the class, using the CC to study calculus, and were not already so strong mathematically that they did not need tutoring. These observations justify dropping the observations with high Cook's D values from the model.

The empirical model is as follows:

(1) $$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 CC_i + \beta_j \text{Control}_{\text{ji}} + \epsilon_i$$

where the model includes j = 1, …, k control variables and i = 1, …, n observations with

(2) $$P_i = \frac{\text{Performance}_\text{i}}{636}$$

Equation 1 is weighted by the variance function $V(P_i) = 636\ P_i\ (1 - P_i)$ to achieve a homoscedastic error term. Performance is being modeled as the sum of 636 independent homogeneous Bernoulli trails, $\text{Performance}_\text{i} \sim \text{Binomial}(636, P_i)$, implying

(3) $$E[\text{Performance}_\text{i}] = \frac{\mathrm{e}^{L_i}}{1+e^{L_i}} * 636$$

To measure the accuracy of the model, we will construct the variable Pass Prediction as

(4) $$\text{Pass Prediction}_\text{i} = \begin{cases} 1, & for\ E[\text{Performance}_\text{i}] > 445 \\ & 0,\ \text{otherwise} \end{cases}$$

and the variable

(5) $$\text{Pass}_\text{i} \ = \begin{cases} 1, & \text{Student}_\text{i} \text{ recieved C or better} \\ & 0,\ \text{otherwise} \end{cases}$$

to estimate the percent correct pass prediction for sampled students as follows:

(6) $$\text{Correct Pass (\%)} = \left(1 - \frac{\sum|\text{Pass Prediction} - \text{Pass}|}{N}\right) * 100$$

We also predict the number of students who, assuming they received the average benefit from visits to the CC and assuming causality, could have passed the class if they had gone to the CC two times per week for 15 weeks and the number of students who passed because of the visits they made to the CC by summing the variables from equations (7) and (8). Let $\hat{P}_i$ denote predicted performance for student `i'.

(7) $$\text{Pass if used optimal CC}_\text{i} = \begin{cases} 1, \hat{P}_i - \beta_1 CC_i + \beta_1(30) > 445 \\ \quad 0,\ \text{otherwise} \end{cases}$$

(8) $$\text{Pass from CC}_\text{i} = \begin{cases} 1, \hat{P}_i - \beta_1 CC_i < 446 \\ \quad 0,\ \text{otherwise} \end{cases}$$

Results

Models of all possible subsets of the test variables along with the required variables were analyzed to determine the final model, where the AICc of the final model is within 2 of the minimum AICc with the fewest independent variables. The final model includes all required variables and all test variables except Total Credits, Age, and Male.

Table 1

Minimum AICc Logistic Regression Results

	All Observations		Cook's D Omission	
	B	Robust SE	B	Robust SE

High School GPA	0.1825***	0.051	0.2034***	0.0436
Midterm 1	0.0132***	0.0026	0.0159***	0.0015
Attendance	0.0274***	0.0033	0.0254***	0.0020
Pretest item: chain rule	0.2363***	0.0512	0.2546***	0.0459
Pretest item: unit conversion	0.1019*	0.0522	0.0940**	0.0431
Pretest item: rate of change	-0.0114	0.0435	-0.0139	0.0363
Low Previous Performance	-0.1112**	0.0492	-0.1581***	0.0409
Visits to CC	0.0063**	0.0026	0.0081***	0.0021
Honors Section	0.0926	0.1077	0.0655	0.0805
# of times taking Calc 1	-0.1434***	0.0366	-0.1605***	0.0303
# of times taking Calc 2	-0.2521***	0.0709	-0.2816***	0.0448
Honors	0.125	0.0857	0.1283*	0.0736
First Generation	-0.06	0.0624	-0.0529	0.0498
Minority	-0.0898	0.0603	-0.0632	0.0469
Masters/Second BA	0.3291**	0.1548	0.3407*	0.2021
International	0.4576***	0.0953	0.5320***	0.0785
Constant	-1.6393***	0.3449	-1.7322***	0.2566
N	683		636	
Deviance/DF	30.0656		20.3085	
Correct Pass	88.1406%		91.3522%	
Passed if Used CC(Total)	68		54	
Passed if Used CC (LLP)	37		34	
Passed from CC Use	37		30	

Note: **p*<.1. ***p*<.05. ****p*<.01 and p-values were not used to select model.

In the more common use of logistic regression with a binary dependent variable, the interpretation of the beta values is generally done by examining the e^{β}, which is the multiplicative change in the odds of an observation being included in the group of interest (commonly labeled as 1) for a one unit increase in the independent variable. With a proportion as the dependent variable, the interpretation is not as straightforward as when inclusion in the group of interest is not the objective. However, the sign of the beta still carries the same general meaning. Negative beta values have an e^{β} less than 1 which is a decrease in odds, while positive beta values have an e^{β} greater than 1 which is an increase in odds. In the context of this example, negative beta values are associated with variables that are believed to decrease the number of course points earned while positive beta values are associated with variables that are believed to increase the number of course points earned.

Using equation (4), a student's expected performance can be calculated using their given characteristics. Using the all observations model, the expected performance of the "average student" in the course, using average values for numerical variables and modal values for categorical variables, is 78.7% with five visits to the CC over the semester (79.3% using Cook's D omission model). The performance of an otherwise similar student who visits the CC twice per week is 81.2% (82.4% using Cook's D omission model). This has a smaller effect on performance than attendance in the course however. The performance of an otherwise average student who attends class half of a standard deviation above average student is 80.5% (81.0% using Cook's D omission model) while the expected performance of an otherwise average student who had half of a standard deviation more visits to the CC than the average student is 79.1% (79.9% using Cook's D omission model).

Knowledge of some pre-requisite material made a marked difference in students' expected performance. Overall, 71.6% of students did not answer the chain rule pretest question correctly. The students that answered the chain rule item correctly have an expected performance score 3.7 percentage points higher for an otherwise average student (for both models). Similarly, the 29.0% of students who correctly answered the gallons to liters unit conversion question have an expected performance score 1.6 percentage points higher (1.5 using Cook's D omission model). A correct answer on the rate of change question was associated with a lower expected performance score, though only .2 percentage points (for both models) and the association is insignificant.

While the endogenous variable Performance ignores points from midterm 1, attendance and extra credit, the model is still able to accurately predict if the sampled students actually passed the class with approximately 88% accuracy for the all observations model and 91% accuracy after omitting outliers. Out of the 683 students included in the first model, we estimate that 37 passing scores may be attributable to the visits these students made to the CC. We also estimate that 68 students could have passed the class if they had gone to the CC two times per week. Thirty-seven of these 68 students reported a C or worse in prior calculus classes, indicating potential success for the CC as an intervention if students with low prior grades were properly targeted. These three estimates should be taken with caution because they unrealistically assume a causal relationship between performance and visits to the CC and assume each student receives the average gain from their visit to the CC. The estimates, when taken as a percent of observations included in model, do not substantially change when omitting outliers, but do slightly decrease.

Conclusion

The results of these analyses suggest that increased visits to the CC is associated with a higher likelihood of passing Calculus 2. In addition to controlling for prior student achievement, as other studies have, this study also includes variables to control for same-semester achievement and motivation by controlling for an early test grade and attendance in the course respectively. In addition, the included independent variables can be used to identify which students are at risk of failing and may be able to pass the course with additional assistance from the CC. This information could be used by teachers to target borderline students and encourage them to seek assistance.

As is common with similar studies, the issue of self-selection is a non-trivial one. Due to this, it is not possible to prove that increased scores are a direct cause of receiving assistance from the CC rather than being caused by other lurking variables such as student motivation, uncontrolled for ability, etc. In a separate survey of the same sample we asked students to respond to the following statement: "I believe that I earned a better grade in the course because of the help at the Calculus Center." Twenty-seven percent of the 151 students who responded strongly agreed with the statement and 29% agreed. Only 10% disagreed or strongly disagreed. Of course, students cannot know for certain the cause of their success in class. However, if most students had said that the CC did not impact their grade, it would be evidence that the correlation we found was primarily due to lurking variables.

Despite our inability to demonstrate causality, we still think the results are significant because of the other studies that found no correlation between visits to a tutoring center and course performance (Cooper, 2010; Walker & Dancy, 2007). Anecdotal evidence suggests that having our tutors attend the course they tutor for and meet weekly with the instructor to discuss the math coming up is one of the factors leading to the CC success.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, vol 19, issue 6, pp. 716-723, doi: 10.1109/TAC.1974.1100705

Annis, L. F. (1983). The processes and effects of peer tutoring. *Journal of Practical Research and Applications*, *2*(1), 39-47.

Arco-Tirado, J. L., Fernández-Martín, F. D., & Fernández-Balboa, J.-M. (2011). The impact of a peer-tutoring program on quality standards in higher education. *Higher Education, 62*, 773-778. doi:10.1007/s10734-011-9419-x

Bamforth, S. E., Robinson, C. R., Croft, A., & Crawford, A. (2007). Retention and progression of engineering students with diverse mathematical backgrounds. *Teaching Mathematics and its Applications, 26*(4), 156-166.

Bressoud, D., Mesa, V., & Rasmussen, C. (2015). Insights and recommendations from the MAA National Study of College Calculus. In D. Bressoud, V. Mesa, & C. Rasmussen (Eds.), *Insights and Recommendation from the MAA National Study of College Calculus*: MAA Press.

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd edition). New York: Springer.

Burns, R. & Burns, R. (2009). *Business Research Methods and Statistics Using SPSS*, chapter 24. SAGE publications. http://www.uk.sagepub.com/burns/website%20material/Chapter%2024%20-%20Logistic%20regression.pdf

Carroll, C., & Gilll, O. (2012). An innovative approach to evaluating the University of Limerick's Mathematics Learning Centre. *Teaching Mathematics and its Applications, 31*(4), 199-214.

Cavanaugh, J. (2009). Model Selection: Lecture II: The Akaike Information Criterion. PowerPoint presentation, The University of Iowa. http://machinelearning102.pbworks.com/w/file/fetch/47699383/ms_lec_2_ho.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2 ed.). NJ: Lawrence Earlbaum Associates.

Colver, M., & Fry, T. (2015). Evidence to Support Peer Tutoring Programs at the Undergraduate Level. *Journal of College Reading and Learning, 46*, 16-41. doi:10.1080/10790195.2015.1075446

Cook, D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics*, Vol. 19, No. 1. Pp. 15-18. https://www.ime.usp.br/~abe/lista/pdfWiH1zqnMHo.pdf

Cooper, E. (2010). Tutoring Center Effectiveness: The Effect of Drop-In Tutoring. *Journal of College Reading and Learning, 40*(2), 21-34. doi:10.1080/10790195.2010.10850328

Cuthbert, R. H., & MacGillivray, H. (2007, 26-30 November). *Investigation of Completion rates of Engineering students.* Paper presented at the Delta '07, the Six Southern Hemisphere Conference on Mathematics and Statistics Teaching and Learning, El Calafate Argentina.

Dowling, D., & Nolan, B. (2006). *Measuring the effectiveness of a maths learning support centre-The Dublin City University experience.* Paper presented at the CETL-MSOR.

Ellis, J., Fosdick, B. K., & Rasmussen, C. (2016). Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: lack of mathematical confidence a potential culprit. *PloS one, 11*(7).

Gujarati, D. & Porter, D., (2009). *Basic Econometrics* 5th edition, ed. Anne Hilbert. 391. New York, NY: Douglas Reiner.

Kyle, J. (2010). Affordability, Adaptability, Approachability, and Sustainability. In C. M. Marr & M. J. Grove (Eds.), *Responding to the Mathematics Problem: The Implemenation of Institutational Support Mechanisms*: MSOR network.

Leung, K. C. (2015). Preliminary Empirical Model of Crucial Determinants of Best Practice for Peer Tutoring on Academic Achievement. *Journal of Educational Psychology, 107*(2), 558-579.

Lidren, D. M., & Meier, S. E. (1991). The effects of minimal and maximal peer tutoring systems on the academic performance of college students. *Psychological Record, 41*(1), 69-78.

Mac an Bhaird, C., & O'Shea, A. (2009). *What type of student avails of mathematics support and why?* Paper presented at the CETL-MSOR Conference.

MacGilliivray, H. L., & Cuthbert, R. H. (2003). *Investigating weaknesses in the underpinning mathematical confidence of first year engineering students.* Paper presented at the Proc. Australasian Engineering Education Conference, The Institution of Engineers, Australia.

MacGillivray, H. L., & Croft, A. (2009). Learning support and students studying mathematics and statistics. *International Journal of Mathematical Education in Science and Technology, 40*(4), 455-472.

Matthews, J., Croft, T., Lawson, D., & Waller, D. (2012). *Evaluation of mathematics support centres: a review of the literature.*

McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models*. 2nd edition. London. Chapman & Hall/CRC

Mills, M. (2017, May, 24th, 2017). [Participants of 2017 Math Learning Resources Leadership Workshop].

Munley, V. G., Garvey, E., & McConnell, M. J. (2010). The effectiveness of peer tutoring on student achievement at the university level. *The American Economic Review, 100*(2), 277-282.

Papke, L. & Wooldridge, J. (1996). Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates. *Journal of Applied Econometrics.* Vol 11, 619-632.

Peduzzi, P., Concato, J., Kemper, E., Holford, T., & Feinstein, A. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*. Vol 49, issue 12, pp. 1373-1379. https://doi.org/10.1016/S0895-4356(96)00236-3, J. P. (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin*, *95*(2), 334.

Pell, G., & Croft, A. (2008). Mathematics Support-support for all? *Teaching Mathematics and its Applications, 27*(4), 167-173.

Perin, D. (2004). Remediation beyond developmental education: the use of learning assistance centers to increase academic preparedness in community colleges. *Community College Journal of Research and Practice, 28,* 559-582. doi:10.1080/10668920490467224

Solomon, Y., Croft, T., & Lawson, D. (2010). Safety in numbers: mathematics support centres and their derivatives as social learning spaces. *Studies in Higher Education, 35*(4), 421-431.

Tallman, M. A., Carlson, M. P., Bressoud, D. M., & Pearson, M. (2016). A characterization of calculus I final exams in US colleges and universities. *International Journal of Research in Undergraduate Mathematics Education, 2*(1), 105-133.

Theobald, R. & Freeman, S. (2013). Is It the Intervention or the Students? Using Linear Regression to Control for Student Characteristics in Undergraduate STEM Education Research. *CBE Life Sci Educ,* vol 13, issue 1, pp. 41-48. doi: 10.1187/cbe-13-07-0136

Thompson, P. W. (2008, 2008). *Conceptual analysis of mathematical ideas: Some spadework at the foundations of mathematics education*.

Thompson, P. W., Byerley, C., & Hatfield, N. (2013). A Conceptual Approach to Calculus Made Possible by Technology. *Computers in the Schools, 30*, 124-147.

Tinto, V. (1998). Colleges as communities: Taking research on student persistence seriously. . *The Review of Higher Education*, *21*(2), 167-177.

Topping, K. J. (1996). The Effectiveness of Peer Tutoring in Further and Higher Education: A Typology and Review of the Literature. *Higher Education, 32*(3), 321-345.

Walker, K. N., & Dancy, M. H. (2007). *Investigaion and evaluation of a physics tutorial center*. Paper presented at the AIP.