

Reflection, Calibration and Achievement In Introductory Calculus

Taylor Kline and Rebecca Dibbs

Texas A&M University-Commerce

Corresponding Author: Rebecca Dibbs Texas A&M University-Commerce, Department of Mathematics, Binnion Hall Room 305, P.O. Box 2011 Commerce, TX 75429; (903) 886-5157; Rebecca.Dibbs@tamuc.edu

Abstract

Studies indicate that calculus acts as a filter for students entering into a STEM discipline, pushing some students to leave their field of study due to failure or disheartenment after going through the course. Since the demand for STEM-trained graduates grows, it is imperative to determine how to prevent more students from switching away from their STEM disciplines. The purpose of this project was to investigate whether the concept of calibration could increase student performance in introductory calculus in order to retain as many students as possible. This was a quasi-experimental study, conducted in calculus I and II courses, imploring CLEAR calculus, which is a nontraditional class style. Results indicated that in calculus I, with data from the first four classroom exams, the calibration could predict performance on the final exam. With calculus II, however, GPA was the only contributing factor to the final exam score. This implies that calibration may be more important in the first course, though further investigation is needed.

Keywords: achievement, ANOVA, calculus, calibration, undergraduate

Introduction

Proteins are very important for the wellness of humans and are the The large numbers of students who switch away from their STEM major during or after taking introductory calculus is a trend that many mathematics researchers are trying to fix. Solving this current issue is important because throughout the United States, and even globally, there is a shortage in the number of workers in the STEM

disciplines (Ellis, Fosdick, & Rasmussen, 2016). One curriculum that shows promise is CLEAR Calculus, an instruction processes developing formalization of calculus concepts that can then be applied to any challenging problem (Oehrtmann, 2008); the course's lab work combined with formative assessment show promise in retaining students; however, we do not understand why this structure helps students stay in STEM past the first semester of calculus (Patterson, 2016). One possible way CLEAR calculus may help students is to improve students' calibration in assessments within the class, which makes students more likely to seek early help. Calibration is based on how accurately a person's self-efficacy matches their execution in a subject area, or what a person expects to achieve versus how they do in terms of a grade (Lin & Zabucky, 1998).

Studying student's calibration was important because of the potential it has to be translated into other classroom settings. Even if the CLEAR Calculus structure was not being used in the curriculum, the calibration aspect could still apply. A quasi-experimental study was designed to test the null hypothesis:

H_0 : There is no relationship between student calibration and performance in introductory calculus.

Literature Review

After reviewing the literature, there were five relevant themes to this research project. They were: why calculus matters, retaining STEM intending students, minority student switchers, curriculum aspects and CLEAR calculus, and calibration. These themes are discussed in order in this section.

Why Calculus Matters

Calculus is the gatekeeper class for STEM majors throughout the world, and students' beliefs that they cannot learn this material is a major obstacle to addressing the shortage of STEM workers (Chiu & Klassen, 2010). It is especially important to teach calculus

effectively to STEM majors not intending to enter mathematics, such as those entering fields of engineering, physics, and STEM education, since the non-mathematics STEM majors are the overwhelming majority in college mathematics courses and the needs of the other STEM departments must be met for all of undergraduate STEM programs to be effective (Rasmussen, Marrongelle, & Borba, 2014). However, there has been little research on how non-traditional calculus curriculums and pedagogy helps these groups of students (Ellis, Kelton, & Rasmussen, 2014), and more work is needed in linking calculus research to theoretical results (Rasmussen et al., 2014).

Retaining STEM Intending Students

Due to the rising shortage of STEM workers globally, researchers are examining the trends for students who remain in their STEM field versus those who switch away from STEM major. One such trend is that calculus tends to turn people away from their originally intended discipline, and this has remained especially true for women (Ellis et al., 2016). Those students who do switch away from calculus have reported that a failed connection between student and teacher is to blame, and, as a result, the student is less likely to seek help on challenging topics (Patterson, 2016). Within the classroom, structured lesson plans that included presenting specific example problems throughout the teaching of new material, having extra material for students to use for practice, holding whole class discussions about topics, asking for students to explain processes and thinking behind problem completion on tests and exams, and lecture-styled instruction helped to contributed to success of retaining STEM intending students (Ellis et al., 2014).

Minority Student Switchers

One goal researchers, employers and educators have is to more thoroughly diversify the STEM career field. The shortage of women is likely due to a lack of confidence rather than a lack of ability, and it is believed that by increasing the number of females who stay in their STEM discipline, the number of women in the workforce would nearly double (Ellis et al., 2016). Certain trends among minority students, including the levels of motivation, preparation

before entering college with class difficulty and standardized testing, levels of support at home and from the family, and financial support, are believed to be contributing factors to student failure in mathematics (Lane, 2016).

Curriculum Aspects and CLEAR Calculus

There are many discussions and arguments on how a calculus course should be structured and taught. One key dispute is the difference in content at the secondary level versus that at the collegiate level, as well as how much focus should be placed on different concepts to benefit other STEM fields (Rasmussen et al., 2014). Other issues include the lack of consistently challenging example problems, the progressive nature of the subject, and the translation of the problem-solving skills learned in calculus into other disciplines (Treisman, 1992). A common theme expressed with difficulties in subject matter comes back to the conceptual understanding of limits in calculus, which is a key, basic topic throughout the course (Oehrtmann, 2008). CLEAR Calculus is a suggested curriculum structure to help to solidify understanding within introductory calculus. This structure allows for students to develop a framework of procedures and steps to transfer to all calculus concepts, creating recognition of identifiable similarities throughout the curriculum of the course (Oehrtmann, 2008). The process includes labs and formative assessment to gauge where students are in understanding the material so that the instructor is aware of what needs to be clarified (Patterson, 2016). It has been found that students who do participate in the formative assessments tend to benefit more and score higher overall (Dibbs, 2015).

Calibration

Self-efficacy and the term calibration go hand-in-hand in educational performance. A student's personal measure of self-efficacy can determine how much effort and time each student will put into their academic achievement (Schunk & Pajares, 2009). Calibration is a metacognitive skill, and it is based on how accurately a student's self-efficacy matches their execution in a subject area, or what they expected to achieve versus how they actually scored (Lin & Zabrocky, 1998). A person with a more positive sense of calibration

tends to express more determination and accomplishment, whereas the opposite tends to be true for those with negative calibration assessments (Chiu & Klassen, 2010).

Although calibration is an important skill for students to master, it is rarely formally taught in mathematics classrooms. However, teachers can help students improve their calibration in several ways. Students should receive feedback on their work to evaluate their personal calibration, in order to ensure that students are learning as much as they possibly can (Lin & Zabrucky, 1998). Students assess their calibration based on their experiences, or lack of experience (Schunk & Pajares, 2009). A student's overconfidence could potentially hinder his or her progress with subject material because they will have an incorrect grasp on key concepts and as a result, will not seek help or tutoring for troubling topics (Champion, 2009). Underestimation of calibration could be equally as destructive to one's achievement because a student may not be motivated to strive to their fullest potential (Chiu & Klassen, 2010).

Methods

The purpose of this study was to investigate the relationship between calibration and final exam performance within a calculus course with the CLEAR Calculus structure.

Study Design

This was a quasi-experimental study. The topic for this study was how a student's sense of calibration could relate to their achieved final exam score. This could not be a true experimental study because students were not randomly assigned to the class. Also, it was unethical to withhold an educational treatment that was more effective in order to have a control.

Study Population & Sample

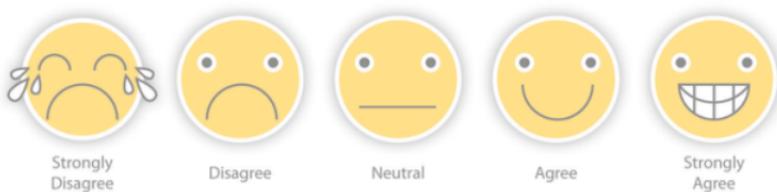
The sample population used in this study consisted of students enrolled in the Fall 2016 calculus I and II courses at a single institution. There were 40 participants total, 20 in Calculus I and 20 in Calculus II, that participated in this study. The students who were entering into the course came from disciplines such as industrial and

construction engineering, computer sciences, physics, and mathematics. The setting of this study was conducted at a rural research university in the southern United States.

The class had the main professor, as well as a TA for Friday classes. The class was held five days a week, on a Monday, Tuesday, Wednesday, Thursday, and Friday schedule. CLEAR Calculus was structured a specific way each week: Monday's were lecture days; Tuesday's were when the lab was held, followed by another lecture day on Wednesday's; Thursday's consisted of lecture, lab and discussion, and Friday's were spent reinforcing the material learned throughout the week. For the purpose of this study, demographic details were collected from the students on the consent forms. Students were asked for their ethnicity, native language, and gender specifications. Based on the data received, each of these were coded with an integer: 0 for male, 1 for female; 0 for English, 1 for Spanish, 2 for Urdu; 0 for Caucasian, 1 for Hispanic, 2 for African American, 3 for Asian, and 4 for Native American. Data collected also included: GPA, actual test scores on the four unit exams and the final, and students' predicted test scores on the four unit exams and the final.

Calibration data was gathered on each test, based on responses on the emoji scale indicated by the students. Students circled an emoji face correlating to their feeling of confidence after completing each problem (see Figure 1).

Figure 1. Affective Likert Scale



Statistical Analysis

Once the data was collected, the emoji scale was converted into numerical values. Then, the test scores and the emoji scores were translated into a calibration score using the methods outlined in Champion (2009). The analysis was conducted using a linear regression. First, the data was checked for a linear relationship by checking to see if there were any significant interaction terms (like calibration*gender) or if there were any terms with a higher R squared value when the data were squared. None were found. Next, the Shapiro-Wilk test was used to check normality; this test revealed that the data was sufficiently normal to do a linear regression.

The initial model that was investigated was:

$$\text{Score} = a_1(\text{calibration}) + a_2(\text{ethnicity}) + a_3(\text{gender}) \\ + a_4(\text{native language}) + a_5(\text{GPA})$$

The initial model was analyzed using each of the demographics and the emoji scores for the four tests as the independent variables, and the final exam graded score as the dependent variable.

Reliability

A Cronbach Alpha statistic to determine the reliability of the instrument was used. Ideally, the Chronbach Alpha would result in at least 0.6, preferably closer to 0.8 (Gall, Gall, & Borg, 2014; Thorndike & Thorndike-Christ, 2013).

Validity

There are several ways to measure validity. Those used in this quantitative study included external validity, internal validity, content validity, and face validity. Each are described in more detail below.

External Validity. The external validity element that this study had was ecological validity. That is, population validity from random sampling does not exist since a classroom setting cannot truly be sampled randomly, but by studying intact classrooms, the environment where treatments naturally occur was preserved (Gall et al., 2014; Thorndike & Thorndike-Christ, 2013).

Internal Validity. For correlational studies, with less direct interaction with participants, the concern was confounding variables (i.e. native language, gender, first generation college students, and student prior knowledge) (Gall et al., 2014; Thorndike & Thorndike-Christ, 2013). The sample was described previously in such a way in order to disclose any potential issues with confounding variables, and they were all included in the initial model.

Content Validity and Face Validity. All tests in the class were constructed using a test blueprint matrix weighted by classroom time spend on each objective. To address face validity, other university staff members including a mathematics education and a mathematician who taught the relevant courses more than ten times examined and approved the instruments being used.

Results

The demographics of the two calculus classes are detailed in Table 1.

Table 1. Summary of participant demographics

| | Calculus I | Calculus II |
|------------------------|------------|-------------|
| Gender | | |
| Female | 4 | 6 |
| Male | 16 | 14 |
| Native Language | | |
| English | 15 | 18 |
| Spanish | 5 | 1 |
| Urdu | | 1 |
| Ethnicity | | |
| Caucasian | 8 | 10 |
| Latino/Latina | 8 | 4 |
| African American | 3 | 4 |
| Asian | 1 | |
| Native American | | 1 |

Calculus I

An initial model for the Calculus I class was created by using SPSS to perform a linear regression model on the initial model. Table 2 provides the results of the initial model.

Table 2. Initial model summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | <i>p</i> |
|-------|------|----------|-------------------|----------------------------|----------|
| 1 | .797 | .635 | .370 | 20.19384 | .000 |

This model was determined to hold significance, because the *p* value was below the significance level of 0.05. Next, an ANOVA test was performed with the data. Table 3 summarizes the results of ANOVA test.

Table 3. Initial ANOVA test

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|-------|------|
| 1 | Regression | 7808.098 | 8 | 976.012 | 2.393 | .090 |
| | Residual | 4485.702 | 11 | 407.791 | | |
| | Total | 12293 | 19 | | | |

Although a *p*-value of 0.090 was not significant, the coefficients were analyzed for any correlations to the final exam score. Table 4 summarizes the coefficients from the initial model.

Table 4. Initial model coefficients

| Model | Unstandardized Coefficients | | Standardized Coefficients | | |
|--------------------|-----------------------------|------------|---------------------------|-------|------|
| | B | Std. Error | Beta | t | Sig. |
| 1 (Constant) | 7.069 | 38.283 | | .185 | .857 |
| Gender | 5.329 | 13.689 | .086 | .389 | .705 |
| Native Language | 6.304 | 11.753 | .110 | .536 | .602 |
| Ethnicity | 3.088 | 6.586 | .106 | .469 | .648 |
| GPA | 4.418 | 13.100 | .116 | .337 | .742 |
| Test 1 Emoji Score | -.023 | .347 | -.022 | -.067 | .947 |
| Test 2 Emoji Score | .053 | .391 | .048 | .135 | .895 |
| Test 3 Emoji Score | .095 | .306 | .098 | .311 | .761 |
| Test 4 Emoji Score | .508 | .398 | .626 | 1.274 | .229 |

With a significance of .229, Test 4 seemed to be accounting the most for the final exam score. Each of the demographics, except for GPA, were removed from the model, as well as all emoji scores besides test four. Table 5 summarizes the final model summary.

Table 5. Final model summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | <i>p</i> |
|-------|------|----------|-------------------|----------------------------|----------|
| 1 | .782 | .661 | .566 | 16.76544 | .002 |

The value of the adjusted R Square increased from the initial model, representing a stronger correlation with the variables, test four emoji score and GPA. Table 6 summarizes the ANOVA test for the final model.

Table 6. Final ANOVA test

| | Model | Sum of Squares | df | Mean Square | F | Sig. |
|---|------------|----------------|----|-------------|--------|------|
| 1 | Regression | 7515.443 | 2 | 3757.721 | 13.369 | .000 |
| | Residual | 4778.357 | 17 | 281.080 | | |
| | Total | 12293.800 | 19 | | | |

The coefficients for the final model are summarized in Table 7. GPA and calibration score on Test 4 contribute approximately equally to the adjusted R-squared value. While the coefficient on GPA was not significant, its large contribution to the explanation of variance warranted retaining this variable in the model. There was not a significant correlation between the GPA and test 4 calibration score ($p = .36$).

Table 7. Final model coefficients

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|-------|------|
| | | B | Std. Error | Beta | t | |
| 1 | (Constant) | 16.161 | 20.466 | | .790 | .441 |
| | GPA | 4.052 | 8.814 | .107 | .460 | .652 |
| | Test 4 | .566 | .188 | .698 | 3.011 | .008 |

Once all of the other independent variables were removed, Test 4 proved to be statistically significant at $p=.008$. The model for this Calculus I course is below. It used GPA and Test 4 calibration to predict the final exam score.

$$\begin{aligned} \text{Final Exam} &= 16.161 + 4.025(\text{GPA}) \\ &+ 0.566(\text{Test 4 Calibration}) \end{aligned}$$

Calculus II

As with the data in Calculus I, an initial model summary for Calculus II was generated by using the same initial model as Calculus I.

Table 8. Initial model summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | <i>p</i> |
|-------|------|----------|-------------------|----------------------------|----------|
| 1 | .558 | .312 | -.189 | 22.00622 | .556 |

The initial model summary for Calculus II showed a low R Square value at 0.312, indicating that there was no significance in the variables. Table 9 summarizes the ANOVA test for the initial model.

Table 9. Initial ANOVA test

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|------|------|
| 1 | Regression | 2413.540 | 8 | 301.693 | .623 | .744 |
| | Residual | 5327.010 | 11 | 484.274 | | |
| | Total | 7740.550 | 19 | | | |

The initial model resulted in a p-value of 0.744, suggesting that there was no correlation between calibration and final exam score. Table 10 summarizes the coefficients for the variables from the initial model.

Table 10. Initial model coefficients

| Model | Unstandardized Coefficients | | Standardized Coefficients | | |
|--------------------|-----------------------------|------------|---------------------------|-------|------|
| | B | Std. Error | Beta | t | Sig. |
| 1 (Constant) | -46.438 | 56.304 | | -.825 | .427 |
| Gender | .143 | 16.298 | .003 | .009 | .993 |
| Native Language | 1.744 | 17.333 | .042 | .101 | .922 |
| Ethnicity | .609 | 5.141 | .036 | .119 | .908 |
| GPA | 14.680 | 11.430 | .397 | 1.284 | .225 |
| Test 1 Emoji Score | .846 | .708 | .444 | 1.195 | .257 |
| Test 2 Emoji Score | -.109 | .267 | -.117 | -.408 | .691 |
| Test 3 Emoji Score | -.117 | .305 | -.122 | -.383 | .709 |
| Test 4 Emoji Score | .171 | .450 | .122 | .381 | .711 |

After initially analyzing the data, none of the predictors significantly accounted for the final exam score. From there, the demographics alone were run with the final exam score, each test calibration score with the final exam score, and GPA with each individual test calibration score with the final exam score. None of the analyses were showing significance for the calculus II final exam score. Throughout the process, GPA had the strongest correlation to the final exam score. Therefore, no model was derived from the collected data in relating student calibration to final exam test score.

Discussion

The question being asked throughout this study was: Is there a relationship between a student's calibration and their performance on the final exam, and can a model be used to predict the final exam

grade based on calibration? For Calculus I, a model was created based on the data. For Calculus II, the result was inconclusive since no significance could be found among the variables.

A model involving a student's GPA and their Test 4 calibration was found to predict that student's final exam grade. With the Calculus I class, calibration from Test 1 had no influence in the final exam score whereas Test 4 had the largest influence. Over the course of the semester, calibration accuracy should increase, and therefore, result in a student applying more time and effort to the outcome of their grade on the final exam (Schunk & Pajares, 2009). This could account for Test 4 mattering the most in the model ($p = .008$). If students have good calibration, pinpointing problem areas would increase efficiency in preparation for tests (Schunk & Pajares, 2009). An unexpected finding that came out of this study was how Test 4 contributed the most to predicting the final exam score. In this Calculus I course, Test 4 covered mostly integral problems, whereas the final exam was mostly comprised of in-depth derivative problems.

There are limitations to this study. One such limitation is the small sample size. Since a convenience sample was used, limited access to participants was the result. Also, while there was diversity in native language and ethnicity among participants, but there was a lack of diversity in gender as very few female students were enrolled. Lastly, the time span of this study may have been a limiting factor, since the study was only conducted over the course of a single semester. Future studies should tracking students who took Calculus I and II with the professor using the CLEAR Calculus curriculum to see how their calibration assessments compared to those in the same Calculus II class who did not have CLEAR Calculus in Calculus I. The results of this study showed that a model for predicting final exam scores of students with the use of calibration could be created for the Calculus I class, but not the Calculus II class. These findings are important for students as well as for teachers. By having a better grasp on their calibration, students can train themselves to prepare for tests accordingly (Schunk & Pajares, 2009). When teachers understand student calibration within their classes, feedback can be

given to students to help them reach a more accurate grasp on their personal calibration (Lin & Zabrocky, 1998). This idea can go beyond calculus and be applied to other subject areas.

References

- Champion, J. (2009). *Self-Efficacy, Calibration, and Exam Performance in College Algebra and Calculus I*. Paper presented at the 12th Conference on Research in Undergraduate Mathematics Education.
- Chiu, M., & Klassen, R. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction, 20*(1), 2-17.
- Dibbs, R. (2015). *Differential participation in formative assessment and achievement in introductory calculus*. Paper presented at the CERME 9th Congress of the European Society for Research in Mathematics Education.
- Ellis, J., Fosdick, B., & 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), e0157447. doi:<https://doi.org/10.1371/journal.pone.0157447>
- Ellis, J., Kelton, M., & Rasmussen, C. (2014). Student perceptions of pedagogy and associated persistence in calculus. *ZDM: The International Journal on Mathematics Education, 46*(4), 661-673.
- Gall, M., Gall, J., & Borg, W. (2014). *Applying Educational Research: How to Read, Do and Use Research to Solve Problems of Practice*. New York, NY: Pearson Higher Ed.
- Lane, T. (2016). Beyond Academic and Social Integration: Understanding the Impact of a STEM Enrichment Program on the Retention and Degree Attainment of Underrepresented Students. *CBE-Life Sciences Education, 15*(3), ar39. doi:10.1187/cbe.16-01-0070
- Lin, L., & Zabrocky, K. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology, 23*(4), 345-391.

- Oehrtmann, M. (2008). Layers of abstraction: Theory and design for the instruction of limit concepts. In M. Carlson & C. Rasmussen (Eds.), *Making the Connection: Research and Teaching in Undergraduate Mathematics Education* (Vol. 73, pp. 65-80). Washington, DC: Mathematical Association of America.
- Patterson, J. (2016). *CLEARing Up Calculus with Adaptive Learning Habits*. (Bachelor of Science), Texas A&M University-Commerce, Commerce, TX.
- Rasmussen, C., Marrongelle, K., & Borba, M. (2014). Research on calculus: what do we know and where do we need to go? *ZDM: The International Journal on Mathematics Education*, 46(4), 507-515.
- Schunk, D., & Pajares, F. (2009). Self-efficacy theory. In K. Wenzel & A. Wigfield (Eds.), *Educational psychology handbook series. Handbook of motivation at school* (pp. 35-53). New York, NY: Routledge/Taylor & Francis Group.
- Thorndike, R., & Thorndike-Christ, T. (2013). *Measurement and evaluation in psychology and education*. New York, NY: Pearson Higher Ed.
- Treisman, U. (1992). Studying Students Studying Calculus: A Look at the Lives of Minority Mathematics Students in College. *The College Mathematics Journal*, 23(5), 362-372. doi:10.2307/2686410