



HAL
open science

Differential participation in formative assessment and achievement in introductory calculus

Rebecca-Anne Dibbs

► **To cite this version:**

Rebecca-Anne Dibbs. Differential participation in formative assessment and achievement in introductory calculus. CERME 9 - Ninth Congress of the European Society for Research in Mathematics Education, Charles University in Prague, Faculty of Education; ERME, Feb 2015, Prague, Czech Republic. pp.2089-2095. hal-01288585

HAL Id: hal-01288585

<https://hal.science/hal-01288585>

Submitted on 15 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Differential participation in formative assessment and achievement in introductory calculus

Rebecca-Anne Dibbs

Texas A&M University-Commerce, Commerce, USA

Prior formative assessment research has shown positive achievement gains when classes using formative assessment are compared to classes that do not. However, little is known about what, if any, benefits of formative assessment occur within a class. The purpose of this study was to investigate the achievement of the students in introductory calculus using formative assessment at the two different participation levels observed in class. Although there was no significant difference on any demographic variable other than gender and no significant difference in any achievement predictive variables between the groups of students at the different participation levels, regular participation in formative assessment was the most significant predictor of achievement in the hierarchical linear model.

Keywords: Approximation framework, calculus, formative assessment, hierarchical linear model, participation.

INTRODUCTION

Introductory calculus is one of the largest choke points for prospective undergraduates who wish to pursue STEM (Science, Technology, Engineering and Mathematics) careers. Students who leave STEM majors are very likely to do so during or immediately after the first semester of calculus (Ellis, Kelton, & Rasmussen, 2014; Rasmussen, Ellis, & Zazkis, 2014). There are several reasons why introductory calculus is a particularly difficult course. The majority of students enrolled in calculus are first-time freshmen, and mathematics and science classes are where students transitioning to higher education are most likely to struggle (Cabrera, Miner, & Milem, 2013; Waterson, Browne, & Carnegie, 2013). Furthermore, the students most likely to leave STEM majors after calculus are from groups that are underrepresented in STEM are-

as: women, first generation college students, English-language-learning students, and students from underfunded urban and rural high schools (Waterson, Browne, & Carnegie, 2013).

One study found that switchers were less likely to feel a sense of connection with their instructors (Ellis, Kelton, & Rasmussen, 2014), which suggests that the increased use of formative assessment, low stakes assignments for instructional planning purposes, may help to increase the number of prospective students in STEM majors past the first semester. The use of formative assessment with undergraduates appears to increase students' perception of a positive relationship with their instructor, make students more likely to seek help, and allows instructors to make data-based decisions on how much review instruction can/should be incorporated into a particular unit (Black & Wiliam, 2009; Dibbs, 2014). For the purposes of this study, formative assessments are defined to be written assignments graded on completion for the purposes of instructor planning.

Regardless of the content area or age of participants, the effect size on most quantitative formative assessment studies is around 0.5 (Karpinski & D'Agostino, 2013). These studies show that classes where formative assessment is used do better on average on common summative assessments than those classes where no formative assessment is used; however, even in classes where formative assessment is used, not all students will regularly complete the formative assignments. The purpose of this study was to investigate the influence of participation on students' growth trajectories on calculus labs designed to develop systematic understanding of limit concepts. Growth can be measured in either student achievement or increases in students' conceptual understanding. The analysis

was delimited to the achievement definition of growth, though qualitative investigations of students' conceptual growth of the approximation framework also showed that regular participants appropriated nearly all limit concepts embedded in the approximation framework while the irregular participants showed little conceptual acquisition beyond procedural fluency (Dibbs, 2014). For this paper, I will distinguish between two different participation levels: regular and irregular. Students regularly participating in the formative assessments missed no more than two of 12 formative assessments during the semester. Although the regular participants earned significantly higher grades on the calculus limit labs, there were students that earned every possible final grade in each participation level.

METHODS

This study is part of a larger QUAL-quan mixed methods case study (Dibbs, 2014). Participants were recruited from two introductory calculus courses taught using the approximation framework at a midsized doctoral granting institution in the Rocky Mountain region. The students enrolled in introductory calculus are most commonly chemistry, science education, mathematics education, or mathematics major, and 35% of the students at the University are first generation college students. There were three sources of data collected: students' assignments, classroom observations, and interviews. The qualitative analysis consisted of daily classroom observations and student interviews. During classroom observations of labs, three of the eight groups in each class were closely observed for peer and instructor interaction: three groups of regular participants, two mixed groups, and one group of irregular participants. During non-lab class days, the instructor's interaction with the class and students' behavior was observed, with particular attention paid to the students observed during labs. After each lab, nine students (both of the mixed participation groups and one regular participation group) were interviewed about their lab write-ups using a cognitive think aloud technique. Students were given a clean copy of their lab and asked to explain each of their answers, who if anyone helped them figure out that particular portion of the lab, and if they would change their answer now; these qualitative data were analyzed to understand students' conceptual growth throughout the semester (Dibbs, 2014). The observations and interviews showed that there was some peer

instruction during the labs, but most of the time, irregular participants only consistently understood the procedural computations portion of the lab. All assignments generated by each participant were collected for the quantitative analysis.

Introductory calculus is a four credit course that met Monday, Tuesday, Wednesday and Friday. The course begins with a brief pre-calculus review and ends with The Fundamental Theorem of Calculus and u-substitution. Students' final grade in the course was determined by online homework (10%), labs (20%), formative assessments (5%), three in-class exams (15% each), and a final exam (20%). Every week followed the same general schedule. On Monday, there was a new section of material introduced and students were given a prelab. Students were asked to complete the Unknown Value portion of the approximation framework (Figure 1) and identify a quantity with which to approximate the unknown value. Students were asked to complete the prelab before class on Tuesday; the prelab was graded on completion at the beginning of class. During class on Tuesday, students worked in groups of three or four on their assigned lab context. After class, students completed a postlab using the online course management software. Each postlab asked students to summarize what their groups did, evaluate how well they understood the material, perform a computation similar to the ones expected on the lab, and identify which portions of the lab they still needed help on. The post-lab was graded on completion, and instructors used students' answers to plan a 20 minute discussion about the lab to begin that class Wednesday.

Although the postlab completion grade was 5% of the students' final grades, the primary purpose of the assignment was to evaluate students' current understanding and plan the next class effectively; in that sense the postlab was primarily a formative assessment. Students were provided automated feedback through the CMS, and it took an average of 15 minutes/week to evaluate a classroom set of postlabs and plan the next class. The remainder of the week was spent on concepts from the textbook. For the derivatives and definite integral labs, the next week would be a repeat of the first; all of the other labs proceeded directly to the regrouping described next. On the third week, students would be placed in new groups, where they were responsible for teaching their context to their new group members; this type of presentation

Fluid traveling at a velocity v across a surface area A produces a flow rate of $F = vA$. Poiseuille’s law says that in a pipe of radius R , the viscosity of a fluid causes the velocity to decrease from a maximum at the center ($r = 0$) to zero at the sides ($r = R$) according to the function $v = v_{\max} (1 - \frac{r^2}{R^2})$. In this activity you will approximate the rate that water flows in a 4-inch diameter pipe if $v_{\max} = 4.44$				
	Contextual	Graphical	Algebraic	Numerical
Unknown Value				
Approximation				
Error				
Error Bound				
Desired Accuracy				

Figure 1: Definite Integral Lab task and approximation framework

is called a Jigsaw presentation because each student is responsible for one piece of a larger idea. After this Jigsaw presentation, students were expected to write up their individual answers to the 20 parts of the approximation framework; this assignment was the summative assessment of each lab. Each lab had one formative prelab and two or three postlabs associate with the summative lab writeup.

The approximation framework is built upon developing systematic reasoning about conceptually accessible approximations and error analyses but mirroring the rigorous structure of formal limit definitions and arguments (Oehrtman, 2008). Approximation is the most common of the seven informal understandings of the definition of a limit; by incorporating the approximation labs into the curriculum, all students are trained to conceptualize limits in the same manner; this makes the transition for formal calculus topics easier for both students and instructors (Oehrtman, 2008).

For each calculus concept, students are asked to identify the unknown value that cannot be solved with algebra, an algebraic technique for approximating the unknown, quantify the error, bound the error, and describe how an approximation can be computed to any desired accuracy. Although the context (and calculus context) changed with each lab, students were asked the same questions on each lab; hence, the labs may be considered repeated assessments on the approximation framework concepts since the process was identical on each lab once the appropriate approximation had been determined. Students were asked to represent these five components of the approximation framework contextually (words and pictures), graphically, algebraically, and numerically. Each lab had three or four different contexts; one of

which was more challenging and intended for students that had seen calculus before. While there is a lab every week, this research was delimited to the approximation framework labs dealing with limits, derivatives, and definite integrals. These labs accounted for 10/14 lab sessions in the semester and contained topics common to all introductory calculus courses. The remaining labs dealt with practice on applications of derivatives: Taylor polynomials, optimization, and related rates. There was also a lab before limits on quantitative reasoning that was not considered due to its low reliability.

The courses were taught at the same time and on the same schedule by two equally experienced instructors. All of the lab questions were scored dichotomously so the inter-rater reliability of the lab write-ups was not a concern. The content validity of the assessments was checked by the course coordinator and an additional expert on the approximation framework. Since the labs were scored dichotomously, KR-20 was used to calculate reliability, and all assessments had reliabilities within acceptable levels (Gall, Gall, & Borg, 2007): the limit, derivative, and definite integral labs had KR-20 values of 0.83, 0.72, and 0.78 respectively.

In addition to participants’ lab write-ups, grade predictive variables and demographic information were collected from each participant. There were no significant differences between the sporadic and non-participation groups on all but one of the demographic or grade predictive variables tested ($p > 0.25$) [1]. Female students were significantly more likely to be regular participants in formative assessment ($p = .03$). Since asynchronous formative assessment, like the ones used in this study, require a greater level of organization and engagement, these assignments tend to slightly favor female students (DiPrete, 2013). Despite

the selection bias inherent in the participation levels, there was no significant difference on any measurement of prior knowledge taken at the beginning of the study. These measures of students' prior knowledge, which all indicated students that chose to participate or not participate in the formative assessments did not have significantly different levels of prior knowledge were not included in the model.

There were 66 students that consented to participate in the study; 13 of the students were removed from the sample because they had prior exposure to the labs that could confound the results. Of the 53 students that were new to the approximation framework labs, only seven had no prior exposure to limit concepts in a prior course, and 27 of the students had AP Calculus [2] in high school. Students needed to have completed at least 10/12 formative assessments to be classified as a regular participant. There were 30 students classified as irregular participants; the remaining 23 students participated regularly in the formative assessments. Only students that completed at least two of the four approximation labs were included in the analysis and are included in the table.

The initial analysis used Bonferroni corrected t-tests found that there were significant differences in the mean number of questions answered correctly by students in each participation level on the three labs, which suggested a hierarchical analysis was most appropriate for the data (Raudenbush, Bryk, Cheong, Congdon, & Toit, 2004). After the null model showed significant differences in the intercept and slope between the two participation classifications, the final model was:

$$SCORE_{ij} = \gamma_{00} + \gamma_{01} * REGULAR_j + \gamma_{10} * LAB_{ij} + \gamma_{11} * REGULAR_j * LAB_{ij} + u_{0j} + r_{ij}$$

The dependant variable is the number of questions a student answered correctly; regular and is a dummy coded variables at the student and time level. Gender, ACT Math score (a standardized exam students take their third year of high school in preparation for applying to college), ethnicity, year in school, native language, major, and the Calculus Readiness Test (a standardized exam administered to students on the second day of classes), did not explain significantly more variance when included, and were all discarded in order to retain the most parsimoniousness model.

RESULTS

There were students that earned every possible final grade in both participation levels, but on the labs, the students that regularly participated in the formative assessments that followed answered more items correctly on the lab write ups than the irregular participants. Table 1 summarizes the results of the t test assuming unequal variances. The p-values have all been multiplied by a factor of three to account for multiple hypothesis tests. Despite the correction, the regular participants in formative assessment have a significantly higher mean score than those not participating regularly. While these two groups were not significantly different on any of the grade-predictive measures available at the beginning of the semester, the students that were irregular participants in the formative postlabs actually had slightly higher average scores than the regular participants.

Given that there appeared to be differences in both the initial level of performance and the rate of change in

	Irregular Mean	Regular Mean	p-value
Limits	8.48	13	0.006
Derivatives (Final)	5.29	16.52	<0.001
Definite Integrals	5.51	17.34	<0.001

Table 1: t-test results for mean number of items answered correctly on each lab by participation level

Random Effect	Standard Deviation	Variance Component	d.f	X ²	p-value
INTRCPT1, u ₀	4.005	16.040	42	189.1522	<0.001
Level-1, r	3.97055	15.765			

Table 2: Null growth model results

score from lab to lab, an analysis that accounted for time was more appropriate to explore this phenomenon further. To confirm growth modelling was the appropriate choice of statistic, a null model [3] was run; this model was significant (Table 2), confirming that multilevel modelling was required; the unexplained variance was 0.504 [4].

The only measurement that resulted in a significant reduction of ICC was participation. When included as a Level-2 variable, the dummy code for regular participation reduced the unexplained variance to 0.363, a reduction of 0.141. The final variance component summary is given in Table 3.

The final estimation of the growth model fixed effects showed that students who regularly participated in the formative prelabs and postlabs were able to answer an average of 5.44 more questions correctly when compared to a student of similar ability who did not regularly participate in the formative assignments (Table 4). The maximum likelihood estimation of the number of significant parameters on the intercept was two, but it was not any of the measures collected as part of the set.

Regular participation in formative assessments also had a significant influence on the rate at which students improved on their lab writeups (Table 5). Given the pronounced ceiling effect of the regular participants' scores, it is not surprising that the slope is relatively small; the regular participants started with relatively high scores and had little room for improvement. However, the near-zero slope value for the non-regular participants does not imply that the irregular participants made no learning gains throughout the semester; rather by the end of the semester these students were able to answer the same questions correctly at the end of the semester on integration that they were able to answer correctly about removable discontinuities in the limits lab.

Although the score on the rewrite after accounting for initial score between the two participation groups was not significantly different ($p=0.0501$), the irregular participants' regression in the subsequent lab indicated that they were not able to apply all of the instructor's feedback in a new context. There is a marked ceiling effect on the final derivative lab and the integration lab for regular participants (Figure 2), which interview data (Dibbs, 2014) indicated was due to nearly complete conceptual acquisition of the limit concepts embedded in the labs.

Random Effect	Standard Deviation	Variance Component	<i>d.f.</i>	χ^2	<i>p</i> -value
INTRCPT1, u_0	2.99581	8.97489	41	124.9583	<0.001
level-1, <i>r</i>	3.96119	15.69102			

Table 3: Growth Model (participation) results

Fixed Effect	Coefficient	Standard error	<i>t</i> -ratio	Approx. <i>d.f.</i>	<i>p</i> -value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	9.504938	0.963904	9.861	41	<0.001
REGULAR, γ_{01}	5.447198	1.139113	4.782	41	<0.001

Table 4: Final estimation of growth model fixed effects

Fixed Effect	Coefficient	Standard error	<i>t</i> -ratio	Approx. <i>d.f.</i>	<i>p</i> -value
For SLOPE, β_1					
INTRCPT2, γ_{10}	0.064472	0.446211	0.144	110	0.885
REGULAR, γ_{11}	1.586417	.521549	3.02	110	0.003

Table 5: Final estimation of growth model slope effects

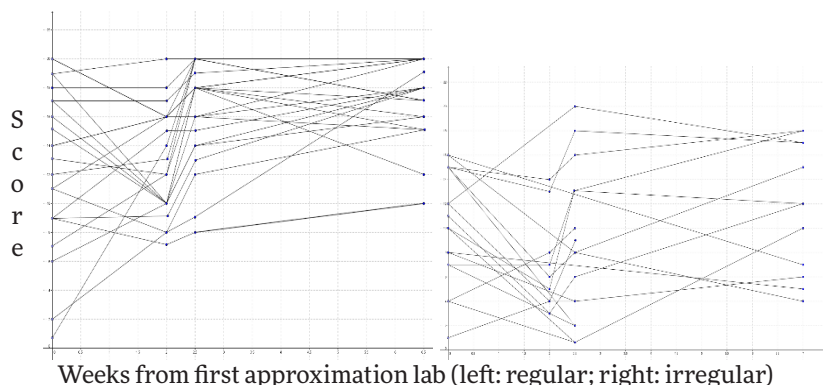


Figure 2: Growth trajectories by participation

DISCUSSION

The results of the study indicate that the students not participating in formative assessments were able to answer fewer questions on average than those students that did participate in the formative assessments, and improved at a slower rate throughout the semester. This is surprising because the students that did not complete any formative assessments attended class for the post-lab based instruction on the day after lab, and the students that did no formative assessments did not have significantly lower levels of prerequisite knowledge than those students participating in the formative prelabs and postlabs. Although this study has a relatively small sample size, this analysis showed that regular participation accounts for 28% of the interclass correlation.

While these results indicated that there were measurable achievement differences between the growth trajectory of those students who participated regularly in formative assessments and those that did not, the analysis also suggested that there were two significant student level factors in the data. Since there were no significant differences on any academic preparation measures for the participants in this study, this suggests that the missing factor in this model is not prerequisite mathematical knowledge. However, given that prerequisite mathematical knowledge is almost always an important factor, further research on the inclusion of this variable or the use of propensity scores in the model is warranted.

One possible explanation is calibration differences between regular and irregular participants in formative assessment. Calibration is considered to be a general metacognitive skill; it is the ability of a learner to accurately assess what they do and do not know (Hacker, Dunlosky, & Graesser, 1998). In this study,

the opportunity for calibration occurred on the limits, first derivative lab, and the definite lab, and it is plausible that the regular participants in the formative assessments are better at identifying the areas in which they need additional help.

There is some support for this supposition in the data. In every lab there was a set of questions that no student asked about on their postlabs. Since none of the students asked for help on the post-lab for these items, an item was considered to be well-calibrated if the student produced the correct solution. In the labs, the statistical evidence for differences in calibration is not clear. The regular participants did answer significantly more of these items correctly, but there were no formal investigations of calibration during this study. Whether pattern of responses is because completing formative assessments on a regular basis helped students maintain a high calibration level or if the formative assessments helped students improve their calibration throughout the semester is an area for future research.

REFERENCES

- Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability*, 21(1), 5–31.
- Cabrera, N. L., Miner, D. D., & Milem, J. F. (2013). Can a summer bridge program impact first-year persistence and performance?: A case study of the New Start Summer Program. *Research in Higher Education*, 54(5), 481–498.
- Dibbs, R.A. (2014). *Formative assessment and students' Zone of Proximal Development in introductory calculus*. (Doctoral dissertation, University of Northern Colorado, Greeley, CO).
- DiPrete, T. A. (2013). *The Rise of Women: The Growing Gender Gap in Education and what It Means for American Schools*. New York, NY: Russell Sage Foundation.

- Ellis, J., Kelton, M. L., & Rasmussen, C. (2014). Student perceptions of pedagogy and associated persistence in calculus. *ZDM*, *46*(4), 661–673.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research* (8th ed.). New York, NY: Pearson.
- Hacker, D. J., Dunlosky, J. E., & Graesser, A. C. (1998). *Metacognition in educational theory and practice*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Karpinski, A. C., & D'Agostino, J. V. (2013). The role of formative assessment in student achievement. In J. Hattie & E. M. Anderman (Eds.), *International Guide to Student Achievement*, (pp. 202–204). New York, NY: Routledge.
- Oehrtman, 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 Practice in Undergraduate Mathematics* (pp. 65–79). Washington, DC: MAA.
- Rasmussen, C., Ellis, J., & Zazkis, D. (2014). Lesson learned from case studies of successful calculus programs at five doctoral degree granting institutions. In *17th Conference on Research in Undergraduate Mathematics Education*. Denver, CO.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & Toit, M. (2004). *HLM 6: Hierarchical Linear and Non-Linear Modeling*. Lincolnwood, IL: Scientific Software International.
- Watterson, C., Browne, W., & Carnegie, D. (2013, August). Steps to increase student engagement and retention in first year engineering. In *Teaching, Assessment and Learning for Engineering (TALE), 2013 IEEE International Conference on* (pp. 1–6). IEEE.

ENDNOTES

1. Age, race, native language, and year in school showed no significant difference on a Chi-Square test. GPA, Math GPA, ACT Math Score, Math GPA, the pre-calculus skills test administered the second day of class, and time elapsed since the previous mathematics course showed no significant differences using Mood's median test.
2. AP Calculus is a one year introductory calculus course generally taught in the last year of high school. The content of this course is the same as the content in the one semester introductory calculus course. Students may elect to take the AP exam at the end of the year. Passing the AP exam (a nationally administered standardized test) would allow students to earn college credit for introductory calculus. Students in

introductory calculus who took AP Calculus in high school did not take/ did not pass this exam.

3. A Null Model assumes that the intercept and slope are both constant for all participants. Failure to reject this test indicates repeated measures ANOVA to be the appropriate test.
4. HLM reports unexplained variance rather than R^2 . The unexplained variance is $1-R^2$.