A Comparison of Traditional and Active Learning Classroom Practices in Calculus Using Fixed- and Mixed-Effects Models

Edgar Fuller	Pablo Duran Oliva	Charity Watson
FIU	VCU	FIU
Adam Castillo	Laird Kramer	Geoff Potvin
UT Arlington	FIU	FIU
	Fulya Eyupoglu	
	FIU	

Active learning in STEM classrooms has been shown to increase student outcomes in multiple ways. We present here a discussion of data analysis supporting these types of conclusions using data from a large-scale randomized study of the implementation of an active learning-based curriculum for Calculus I that collected 1019 observations of student outcomes over three semesters. We compare fixed-effects models including cluster levels of student learning outcomes to mixed-effects models with random cluster level effects. We discuss the differences between the models and the resulting effect sizes suggested by the different factors included in the models.

Keywords: calculus, regression model, fixed effects, mixed effects

Introduction

Evidence of the effectiveness of active learning (AL) classrooms in STEM includes increases in student achievement, attitudes, and persistence in college (Duran et al., 2022; Ellis et al., 2016; Freeman et al., 2014; Kramer, Fuller, Watson, Castillo, Oliva, et al., 2023). Most of these studies have measured this effectiveness by conducting non-randomized experiments and focused on evidence gathered from observational studies or performed meta-analyses of studies with varying implementations in order to interpret the outcomes across a collection of studies with different sets of participants or contexts. In (Kramer, Fuller, Watson, Castillo, Oliva, et al., 2023), the authors present a large-scale study of the implementation of an active learning-based curriculum for Calculus I that collected 1019 observations of student outcomes over three semesters with a variety of instructors teaching at multiple times. On the one hand, this study presents compelling evidence that student learning in an active curriculum has the potential to be far more effective than in traditional lecture-based paradigms. The random allocation of students to the treatment and control conditions allows the comparison of student outcomes across populations while controlling for multiple demographic factors including incoming mathematics background, race and/or ethnicity, gender, and student enrollment status. At the same time, students are grouped into sections at the same time of day/day of week and instructors will be present in different sections, introducing random effects on student outcomes (Hedges, 2009).

In this work, we present a fixed effects model (FEM) of student learning outcomes dependent on multiple demographic factors and show that the differences between the treatment and control group outcomes are statistically significant. We then compare that FEM to the mixed-effects model (MEM) in (Kramer, Fuller, Watson, Castillo, Oliva, et al., 2023) and discuss the ways in which the random effects in that model compare to the FEM properties. We also discuss changes to the MEM and show that the MEM presented in (Kramer, Fuller, Watson, Castillo, Oliva, et al., 2023) and using the data set from that project here (Kramer, Fuller, Watson, Castillo, Duran

Oliva, et al., 2023) possesses the same structure as both the FEM as well as these other models. Finally, we provide a comparison of the effect sizes and confidence intervals for different models.

Methods

In (Castillo et al., 2022; Duran et al., 2022; Kramer, Fuller, Watson, Castillo, Oliva, et al., 2023) the authors collected data from a trial of an active learning based curriculum where students were randomly allocated and presented a number of outcomes analyses. In this study, 1019 students were randomly allocated to either the treatment condition which utilized the Modeling Practices in Calculus (MPC) approach to student learning or the control setting based on existing instructional practices that were primarily lectured based. Student data was collected from institutional records that provided potential fixed demographic effects on student outcomes including mathematics background from standardized tests/high school GPA (*MBS*), gender (*Gender*), race and/or ethnicity (*RE*). Along with these data, the section of the course (*Section*) and the instructor of record (*TIDw*) were recorded as random effects that depended on student choices made during enrollment. In addition, student outcomes were recorded both as grade outcomes in the course as well as a measure of student learning, *LearningMeasure*, collected from embedded final exam questions implemented within the common final administered across all sections of the courses and assessed in a blinded form by multiple evaluators (Kramer, Fuller, Watson, Castillo, Oliva, et al., 2023).

Using these data, a fixed effects model with interacting terms can be constructed to determine the dependence of the measure of student learning from the study, *LearningMeasure*, on these factors including the cluster levels *SecPair* to represent the grouping of participants by their time of day and day of week classroom groups. Note that these sixteen pairs then split into 32 based on random assignment, and this is also a representation of the instructor level clustering. The linear model used here is then

$$\begin{aligned} LearningMeasure_i \sim & \beta_0 + \beta_{i1} Treatment_i + \beta_{i2} MBS_i + \beta_{i3} Gender_i + \beta_{i4} RE_i + \\ & \beta_{i5} SecPair + \\ & \beta_{jkl} Treatment_{ij} MBS_{ik} Gender_{il} + \epsilon_i \end{aligned}$$

where *LearningMeasure* is the vector of students' scores, *Treatment* is the predictor (a dummy variable, 0:Non-MPC sections; 1: MPC sections), β is the regression coefficient for the covariates measured, and ϵ is the vector of residuals, assumed to be distributed N(0,2I). Other variations involving either *Section* or *TIDw* for cluster levels as fixed effects were not considered, as they introduced collinearity with *Treatment*. Similarly, a variation of the mixed effects model presented in (Kramer, Fuller, Watson, Castillo, Oliva, et al., 2023) can be constructed using the fixed effects above but instead applying *SecPair* along with *Section* and *TIDw* as random effects. Models were implemented in R (R Core Team, n.d.) using the glm and lme4 (Kuznetsova et al., 2017) packages.

Results

Coefficients for the FEM with both interacting and non-interacting terms along with a model with Treatment as the only fixed effect were computed and found to have the values shown in Table 1. The model shown using the sixteen cluster levels of *SecPair* also has the interaction terms of the first model.

Table 1. Fixed Effects Model Coefficients with and without Interactions and Cluster Levels.

	Model With Interactions			**No Interactions**			**With SecPair**		
Predictors	*Est.*	*s.e*	*p*	*Est.*	*s.e*	*p*	*Est.*	*s.e*	*p*
(Intercept)	-8.20	9.06	0.365	-2.99	4.91	0.543	-0.54	9.41	0.954
Treatment	30.48	11.82	**0.010**	16.45	1.39	**<0.001**	28.91	11.63	**0.013**
MBS	0.89	0.13	**<0.001**	0.81	0.06	**<0.001**	0.84	0.13	**<0.001**
Gender	4.79	12.49	0.701	-3.64	1.39	**0.009**	0.25	12.24	0.984
RE2	-7.68	3.59	**0.032**	-7.57	3.58	**0.035**	-9.66	3.56	**0.007**
RE3	-1.04	2.64	0.694	-1.17	2.64	0.656	-2.30	2.60	0.376
RE4	2.21	3.38	0.512	2.43	3.37	0.471	-1.12	3.35	0.737
Treatment × MBS	-0.22	0.18	0.218				-0.19	0.17	0.287
Treatment × Gender	-23.97	16.74	0.152				-20.32	16.38	0.215
MBS × Gender	-0.13	0.18	0.479				-0.07	0.18	0.707
(Treatment × MBS) × Gender	0.37	0.25	0.136				0.30	0.24	0.213
SecPair2							-0.74	3.62	0.839
SecPair3							2.55	3.61	0.481
SecPair4							-12.79	3.90	**0.001**
SecPair5							-6.54	3.79	0.084
SecPair6							-0.93	3.65	0.799
SecPair7							4.97	4.23	0.240
SecPair8							-16.92	4.61	**<0.001**
SecPair9							-0.59	3.57	0.868
SecPair10							-3.00	3.43	0.382
SecPair11							2.19	3.54	0.536
SecPair12							-9.44	3.59	**0.008**
SecPair13							-5.19	3.37	0.123
SecPair14							2.24	3.78	0.553
SecPair15							-1.07	4.68	0.819
SecPair16							-4.27	4.97	0.390
Obs	671			671			671		
R^2	0.331			0.328			0.384		
AIC	5792.7			5787.4			5767.6		
logLik	-2884.3			-2885.7			-2856.8		

Using these FEMs we can also then estimate the effect sizes for the fixed effects, shown in Table 2.

Table 2. Effect Sizes for Fixed Effects in Interacting Model

Parameter	Partial Cohens *f*	95% CI low	95% CI high
Treatment	0.45807	0.37679	0.539
MBS	0.54491	0.46202	0.627
Gender	0.10635	0.0287	0.184
RE	0.12148	0.00737	0.189
SecPair	0.29389	0.16633	0.338
Treatment:MBS	0.00917	0	0.079
Treatment:Gender	0.00445	0	0.063
MBS:Gender	0.03243	0	0.11
Treatment:MBS:Gender	0.049	0	0.126

Computing a mixed effects model incorporating interactions and random effects due to section (*Section, SecPair*) and teacher level (*TIDw*) as in (Kramer, Fuller, Watson, Castillo, Oliva, et al., 2023) yields the coefficients shown in Table 3.

Table 3. Mixed Effects Model Coefficients with and without Interactions.

	Model With Interactions			**No Interactions**		
Predictors	*Estimates*	*std. Error*	*p*	*Estimates*	*std. Error*	*p*
(Intercept)	-2.80	8.98	0.755	0.67	5.14	0.896
Treatment [TR]	28.07	11.71	**0.017**	15.76	2.78	**<0.001**
MBS	0.83	0.13	**<0.001**	0.78	0.06	**<0.001**
Gender [F]	0.09	12.02	0.994	-3.72	1.35	**0.006**
RE [2]	-9.42	3.49	**0.007**	-9.30	3.48	**0.008**
RE [3]	-2.18	2.55	0.395	-2.26	2.55	0.376
RE [4]	-0.51	3.28	0.877	-0.29	3.27	0.930
Treatment × MBS	-0.18	0.17	0.285			
Treatment × Gender	-18.83	16.09	0.242			
MBS × Gender	-0.05	0.18	0.769			
(Treatment × MBS) × Gender	0.28	0.24	0.247			
Random Effects						
σ^2	289.50			288.68		
τ_{00}	17.14 $_{\text{Section}}$			17.46 $_{\text{Section}}$		
	13.67 $_{\text{TIDw}}$			13.57 $_{\text{TIDw}}$		
	12.62 $_{\text{SecPair}}$			12.45 $_{\text{SecPair}}$		
ICC	0.13			0.13		
N	16 $_{\text{SecPair}}$			16 $_{\text{SecPair}}$		
	32 $_{\text{Section}}$			32 $_{\text{Section}}$		

	19 TIDw	19 TIDw
Observations	671	671
Marginal R^2 / Conditional R^2	0.304 / 0.395	0.303 / 0.394
AIC	5754.594	5746.831
log-Likelihood	-2862.297	-2862.415

Computing effect sizes within a MEM is more complicated and in general must be approximated (Hedges, 2009). The effect sizes for each section pair were computed and shown in Table 4, and the precision weighted effect size (DerSimonian & Laird, 1986) for the Treatment effect computed from those using the standard error for each.

Table 4. Effect Sizes for Each Section Level Pair of Treatment and Control groups

		95% Confidence Interval	
Section Pair	Hedges *g*	lower	upper
1	1.097	0.483	1.748
2	-0.501	-1.094	0.076
3	1.003	0.400	1.639
4	2.221	1.412	3.127
5	-0.259	-0.903	0.375
6	0.569	-0.001	1.158
7	0.200	-0.591	1.004
8	1.662	0.655	2.800
9	0.673	0.104	1.262
10	0.993	0.453	1.560
11	0.932	0.362	1.530
12	0.899	0.309	1.518
13	1.226	0.699	1.782
14	1.021	0.375	1.706
15	0.740	-0.185	1.727
16	0.808	-0.142	1.828

In some cases, the variance of the outcome data due to the number of data points in some pairs increases the error and uncertainty in the outcomes for that cluster. This pattern is reflected in a forest plot of the effect sizes (Hedges' *g*) computed within the section pair clusters as shown in Figure 1.

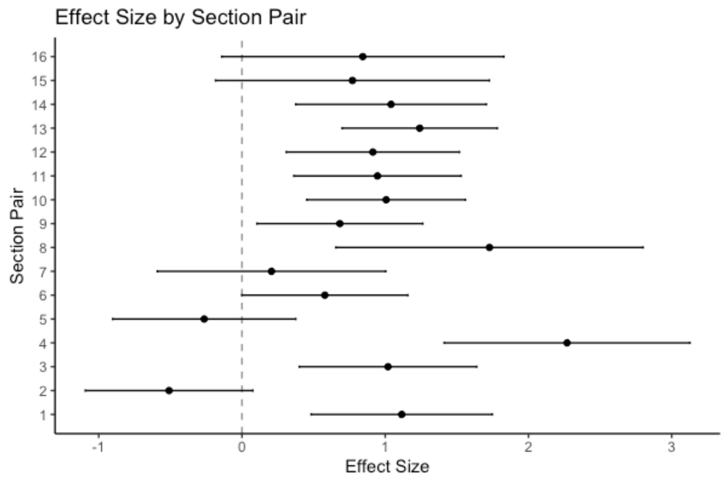

Figure 1. Forest Plot of Section Pair Effect Sizes

Of the 16 pairs, twelve have statistically significant increases in outcome for the Treatment section, and four do not. Two of the four with indeterminate outcomes are positive effects, and two are negative. The precision weighted effect size in this case is d_{pw}= 0.8008 (CI [0.4589, 1.1427], t=4.99 p=0.0002).

Discussion

The fixed effects model indicates that the treatment condition is statistically significant even in the presence of the other fixed effects. Specifically, the models show that the treatment is statistically significant and positive even in the model with interactions (β = 30.48, 95% CI [7.30, 53.65], t(660) = 2.58, p = 0.010) when controlling for mathematics background, race/ethnicity, and gender. The FEM interaction terms are found to be insignificant (the null hypothesis of zero-valued coefficients cannot be rejected). In addition, the model without interactions is found to have a better fit (AIC = 5787.4) than the model with interactions (AIC = 5792.7), but the model with interactions and the SecPair cluster levels is best (AIC = 5767.6) with a coefficient of β =28.91 which is almost equivalent to the non-interacting estimate. This model compares well to the mixed effects model that includes random effects from the *SecPair*, *Section* and *TIDw* clusters. In that model, fixed effects were found to explain 39.4% (with marginal R^2=0.303) of the outcome while the FEM was found to explain 38.4% (R^2 = 0.384) of the outcome with interactions and SecPair cluster levels. The MEM is perhaps preferred since it incorporates the impact of student choice of section and the presence of the instructor, but this analysis shows that these factors explain approximately ~10% (conditional R^2=0.1) of the outcome in this data. In determining the treatment effect on outcomes, the FEM gives a similar model for interpretation even though the MEM is a slightly better fit. One benefit of the MEM approach, however, is that it also characterizes the intraclass correlation (ICC = 0.13) for these underlying clusters and provides some insight into the degree to which those correlations might impact the outcomes. In (Kramer, Fuller, Watson, Castillo, Oliva, et al., 2023), the fixed effect Cohen's d = 0.774 (CI [0.618,0.930]), and cluster adjusted effect size from the random effects of the MEM d_T =0.771*,*CI [0.468,1.073] are consistent with the metanalytic approach here using only the section pair data.

The coefficient similarities between the linear FEM and the MEM with random effects in some sense reflect the fact that the data form distinctly linear structures when modeled on the strongest effect, *MBS*. As shown in Figure 2, a LOESS curve approximation of the data plotted within a scatterplot of *LearningMeasure* against *MBS*, we see that the majority of the curve is linear in nature with low levels of uncertainty, and that only in the outlier regions where few data points exist do the curves show high levels of non-linearity and uncertainty.

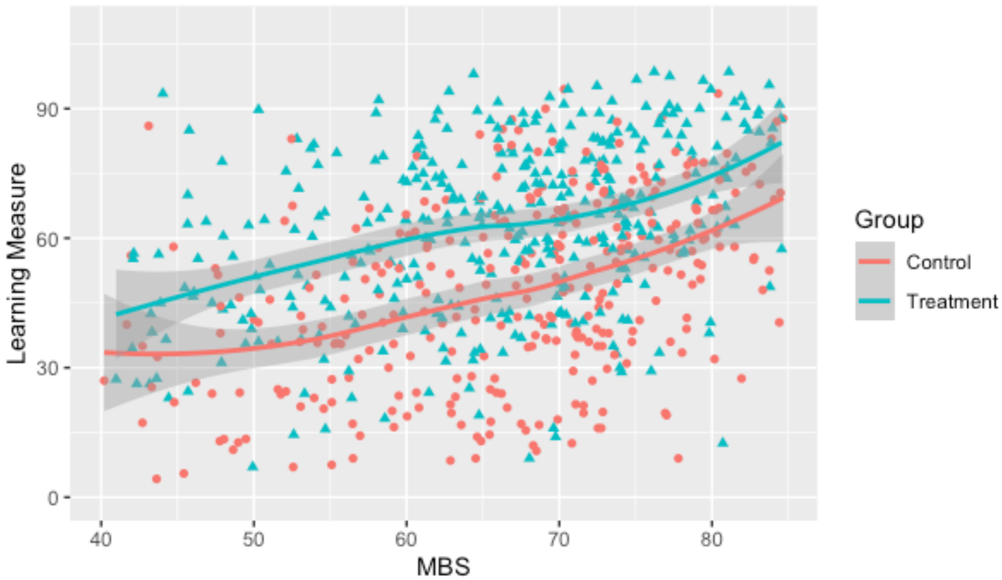

Figure 2. Scatterplot of LearningOutcome Against MBS (Math Background Score) for Treatment and Control Groups with LOESS Curve Approximations. Dark Grey Areas Indicate the 95% CI Surrounding the Approximating Curve

Conclusions

In the data analyzed in this article, the underlying cluster structure has a relationship to the outcome variable variance but this explanatory power is small when compared to the other fixed and random effects. The impact of the time of day/day of week of the courses was on the order of 50% of the effect found for *Treatment* and *MBS* in the FEM. The estimates for the coefficients of *Treatment* in all the models are similar and the computed effect sizes are consistent using either model. The MEM approach has the advantage of estimating the relative variance due to cluster levels in a way that represents their underlying random nature, while incorporating *SecPair* into the FEM approach yields similar results. Both analyses support the conclusion that the treatment condition led to medium to large increases in student learning outcomes.

Acknowledgments

We thank the National Science Foundation for supporting this work under DUE # *1832450*.

References

Castillo, A., Duran, P., Fuller, E., Watson, C., Potvin, G., & Kramer, L. H. (2022). Student attitudes in an innovative active learning approach in calculus. *International Journal of Mathematical Education in Science and Technology*, 1–29. https://doi.org/10.1080/0020739X.2021.2023771

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188.

Duran, P. A., Castillo, A. J., Watson, C., Fuller, E., Potvin, G., & Kramer, L. H. (2022). Student attitudes and achievement in active learning calculus. *International Journal of Mathematical Education in Science and Technology*, 1–22.
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), e0157447. https://doi.org/10.1371/journal.pone.0157447
Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111
Hedges, L. V. (2009). Effect sizes in nested designs. *The Handbook of Research Synthesis and Meta-Analysis*, *2*, 399–416.
Kramer, L., Fuller, E., Watson, C., Castillo, A., Duran Oliva, P., & Potvin, G. (2023). *Establishing a New Standard of Care for Calculus Using Trials with Randomized Student Allocation* (10.5061/dryad.kkwh70s95; v3.0) [dataset]. Dryad; Dryad. https://doi.org/10.5061/dryad.kkwh70s95
Kramer, L., Fuller, E., Watson, C., Castillo, A., Oliva, P. D., & Potvin, G. (2023). Establishing a new standard of care for calculus using trials with randomized student allocation. *Science*, *381*(6661), 995–998. https://doi.org/10.1126/science.ade9803
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/doi:10.18637/jss.v082.i13
R Core Team. (n.d.). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org