The Relationship Between Course Scheduling and Student Performance

Seth Poulsen
University of Illinois at Urbana-Champaign
sethp3@illinois.edu

Carolyn J. Anderson
University of Illinois at Urbana-Champaign
cja@illinois.edu

Matthew West
University of Illinois at Urbana-Champaign
mwest@illinois.edu

ABSTRACT

Using 10 years of grade data from a university computer science department we fit a multi-level proportional odds model and find that students earn a higher grade in an afternoon class at 1.15 times the odds for a morning class, even when controlling for GPA. This finding has implications both for student learning and for experimental studies that compare classes without considering the time of day at which they are taught. We find that there are no significant trends for student performance based on term when looking at the department as a whole, though there are such trends for certain courses in particular.

Keywords

course scheduling, GPA, research methods, multi-level models

1. INTRODUCTION

When evaluating the effectiveness of a new instructional technique or educational intervention, researchers would ideally test the intervention on many sections of the same course to increase confidence that the intervention is working as intended. Data would then be analyzed using multi-level regression to properly treat the variance that naturally happens between sections of the same course [24].

Multi-levels models have been used in computer science education, for example, [21], but they are few and far between. Even thought there have been many multi-institution, multinational studies in computer science, [2, 6, 8, 9, 10, 18, 23] even they often don't include enough clusters of data to be able to use a multi-level model.

As researchers and educators we understand the reasons that larger studies are not undertaken more often: even planning an educational intervention experiment with one experimental and one control section can be very resource intensive! In many situations, especially when first piloting new educational techniques, it is completely impractical to expect that researchers will be able to experiment on more than a single section of a course.

Unfortunately, experimenting with one or only a few sections of a course requires the researcher to make the assumption that essentially all things are equal about the students taking the courses and the courses themselves, apart from the intervention.

Despite controlling for as many factors as possible, such as instructor, course assignments, tests, and more, there are still often factors that lie outside the researcher's control, such as the day of the week, time of day, term, and location that their course is scheduled for. Furthermore, students self-select into which section of the course that they want to take! These variations between sections may be introducing a selection bias threatening the validity of these educational experiments.

In this year's SIGCSE Technical Symposium alone, there were 7 studies which tested new educational practice through experimenting with either one or just a few experimental and control sections of the same course, operating either explicitly or implicitly under the "all things equal" assumption [7, 12, 13, 15, 19, 20, 25]. While six of these seven studies clearly stated the year and term of the sections that they collected data from, only one of them stated the time of day and days of the week on which the sections were held. This method of comparing one or only a few sections of a course in assessing instructional practice is also used in other areas of discipline-based education research, including chemistry [26], physics [14], and materials science [16], to name a few.

The desire to check the validity of the "all things equal" assumption for experimenting with multiple sections of the same course, and discussion with colleagues, led us to the following research questions:

1. Is student performance in a course related to the time of day the course is scheduled for?
2. Is student performance in a course related to the term?
3. Is student performance in a course related to the days of the week the course is held on?
4. Is student performance in a course related to the building in which the course is held?

2. LITERATURE REVIEW

It is has been shown that adolescents struggle to perform to their fullest potential early in the morning, causing many school districts to push back school start times [5, 11]. However, there has not been enough work done to verify that this effect also holds true for college students [17].

Marbouti et. al. analyzed data from 15 different sections of an introductory university engineering course and found that due to lower attendance in morning sections, the early morning sections of the course significantly under-performed other sections [17]. To our knowledge, no one so far has examined a data set including more than one course to see if this trend holds generally.

Most literature agrees that courses offered in condensed terms (such as most universities' summer terms) lead to the students learning the material equally well or even better than courses that are taught over a full length term [1].

When it comes to day-of-week scheduling, there is quite a division in the literature, with some studies finding that spacing lessons out over the week more helps students learn more, while others find that students perform just as well when the course material is presented only one day a week. [4]. Some studies even suggest that the outcome depends on whether the material requires deep comprehension and analysis, or simply recall [4].

3. DATA

Our grade and course scheduling data was acquired from the registrar at the University of Illinois at Urbana-Champaign. Because our primary focus is the relationship between course scheduling and time of day, we removed topics and reading courses that were only taught once, as well as courses that are not scheduled, such as independent study and senior thesis courses. Summer courses were also removed from the data set, to avoid comparing versions of the same course which were taught on an entirely different time scale, and sometimes even with a different set of instructor expectations.

Drops and withdrawals were also removed from the data set. After cleaning the data, we were left with 72,739 student grades from 24,705 students across 1,938 sections of 101 courses. The grade data consists of letter grades, which we converted to grade points for the purposes of fitting the model (A $\rightarrow$ 4.0, A- $\rightarrow$ 3.67, B+ $\rightarrow$ 3.33, etc.). The overall mean grade in the data set is 3.100, and the median is 3.33 (B+).

At the University of Illinois, the Fall term starts in late August and ends in mid-December, and the Spring term starts in mid-January and ends in mid-May. We chose from the beginning to treat time of day as a categorical variable, where courses beginning before 10:00 a.m. were considered "Morning," courses starting between 10:00 a.m. and 2:00 p.m. were considered "Midday," courses starting between 2:00 p.m. and 5:00 p.m. were considered "Afternoon," and courses which started after 5:00 p.m. were considered "Evening" courses.

A look at the data set shows that of the 101 courses offered in the computer science department, 54 of them were always

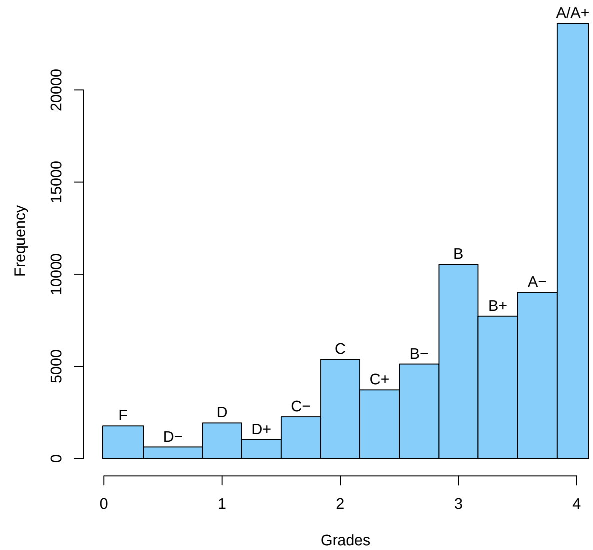

Figure 1: All student grades in the data set.

taught on the same days of the week every time they were taught, and another 30 were only taught on 2 different day configurations (e.g. a class was taught either Monday and Wednesday or Tuesday and Thursday, but not in any other day configurations). Because of this, we chose to leave day of the week considerations out of our analysis entirely.

Figure 1 shows the distribution of student grades from the entire data set. As hinted at by Figure 2 and Figure 3, and revealed by deeper data exploration, there was sufficient variance in the performance of students between courses as well as between sections of the same course to justify grouping the data by course and by section for the fitting of the model, leading to a three-level model.

4. METHODS

We fit the data using a three level model of the following form, where students are indexed by i, sections are indexed by j, and coursed are indexed by k, and y represents some grade (e.g. A, A-, B, etc.):

- Level 1 (student):

$$\ln\left(\frac{P(\text{grade}_{ijk} < y)}{P(\text{grade}_{ijk} \geq y)}\right) = \beta_{0jk} + \beta_{1jk}\text{GPA}_{ijk}$$

- Level 2 (section):

$$\beta_{0jk} = \gamma_{0k} + \gamma_{1k}\text{Midday}_{jk} + \gamma_{2k}\text{Afternoon}_{jk} + \gamma_{3k}\text{Evening}_{jk} + U_{jk}$$

- Level 3 (course):

$$\gamma_{0k} = \delta_0 + W_k$$

Midday_{jk}, Afternoon_{jk}, and Evening_{jk} are dummy codes denoting the time of day a course was held, and all of them

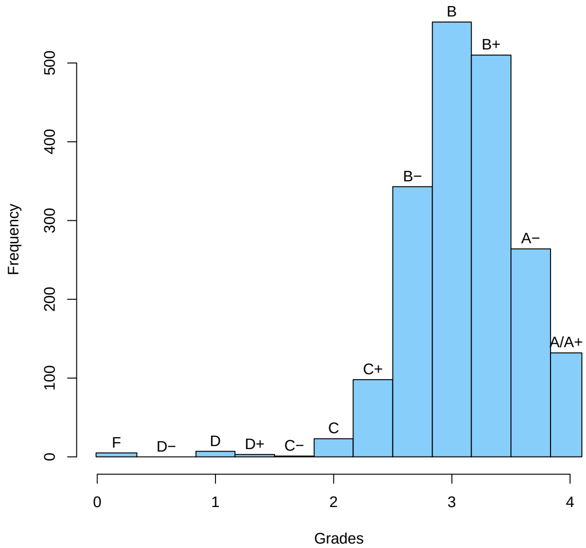

Figure 2: Distribution of section averages

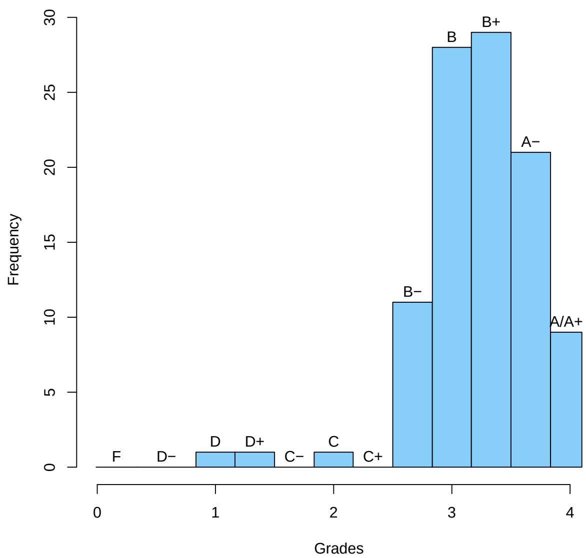

Figure 3: Distribution of course averages

being 0 represents a Morning class. Our model assumes that the error term on the section level, U_{jk}, and the error term at the course level, W_k, are multivariate normal distributions which are independent of one another. After substituting and gathering the error terms, we obtain the mixed model:

$$\ln\left(\frac{P(\text{grade}_{ijk} < y)}{P(\text{grade}_{ijk} \geq y)}\right) = \delta_0 + \beta_{1jk}\text{GPA}_{ijk} + \gamma_{1k}\text{Midday}_{jk} + \gamma_{2k}\text{Afternoon}_{jk} + \gamma_{3k}\text{Evening}_{jk} + W_k + U_{jk}$$

We used the R (version 3.6.3) package brms [3, 22] to fit the model using Bayesian estimation. We used brms with default priors, 3 chains, 1500 warm-ups, and 3000 iterations. After examining the $\hat{R}$ values and trace plots, we concluded that the model converged.

We also fit a version of the model including a dummy code for term (Fall vs. Spring) and found that there was no significant general trend for the relationship between term and course. However, fitting similar models for some individual courses revealed that some courses do have significant differences in performance between semesters, with some courses having better performance in the Fall and some having better performance in the Spring.

Finally, we fit a version of the model including a dummy code for whether or not the section was held in the computer science department building, with the hypothesis that sections held in the computer science department building would be more desirable and would thus fill up with more responsible students who registered on time. We found no significant relationship between student performance and which building the course was held in.

5. RESULTS

The estimated parameters of the final model are listed in Table 1. This model allows us to estimate the relative probability that a student will receive each letter grade, given their cumulative GPA, the course and section of the course, and the time of day that the course was scheduled. The probability of a higher grade increases for higher values of GPA. The later in day, the probability of a higher grade increases. According to the model, the odds that a student receives a higher grade in an afternoon class (based on model fit information from Table 1) are $e^{0.14} = 1.15$ times the odds for a student with the same GPA taking the same class in the morning. Additionally, holding all other variable constant, the odds of a higher grade in an evening class are $e^{0.17} = 1.19$ times the odds in a morning class.

To allow an interpretation of the effect size in grade units rather than only as probabilities, we used the model to simulate what the average grade over all data points would be if all courses in the department were offered at the same time of day. The results, shown in Table 2, show that students perform 0.04 and 0.05 grade points better in afternoon and evening classes, respectively, than they do in morning classes. Figure 4 helps us to visualize that student performance is actually monotonically increasing throughout the

	Estimate	Est. Error	Lower-95% Credible Interval	Upper-95% Credible Interval	Significance
Population-Level Effects:					
Intercept (D-)	4.68	0.15	4.40	4.97	***
Intercept (D)	5.09	0.15	4.81	5.38	***
Intercept (D+)	5.94	0.15	5.66	6.23	***
Intercept (C-)	6.27	0.15	5.98	6.56	***
Intercept (C)	6.84	0.15	6.56	7.14	***
Intercept (C+)	7.83	0.15	7.54	8.12	***
Intercept (B-)	8.35	0.15	8.07	8.65	***
Intercept (B)	8.98	0.15	8.69	9.28	***
Intercept (B+)	10.07	0.15	9.78	10.37	***
Intercept (A-)	10.81	0.15	10.53	11.12	***
Intercept (A/A+)	11.70	0.15	11.41	12.01	***
Cumulative GPA	3.26	0.02	3.23	3.30	***
Afternoon	0.14	0.06	0.02	0.26	*
Evening	0.17	0.11	-0.05	0.39	
Midday	0.06	0.06	-0.06	0.17	
Group-Level Effects:					
sd(Course Intercept)	1.15	0.11	0.95	1.37	
sd(Section Intercept)	0.61	0.01	0.58	0.64	

Significance codes: 0 < *** < 0.001 < ** < 0.01 < * < 0.05

Table 1: Estimated Paramaters of the Proportional Odds Model.

Time	Average Grade	Difference from Morning
Morning	3.075	-
Midday	3.092	0.018
Afternoon	3.118	0.043
Evening	3.129	0.054

Table 2: Simulated average grade using the model, if all classes were offered at the same time of day.

day, with the worst performance in morning classes, and the best performance in evening classes.

It is also important to note that there is a large variance in grades between sections and courses, so in addition to the general trends, it appears there is often large variation between any two sections of a given course. In Figure 5, we visualize the relative amount of uncertainty at the section (U_{jk}) and course (W_k) levels. It appears that much more variance in course performance comes from the course rather than the section level, but there is still a significant amount of unexplained variance between sections in our model.

6. DISCUSSION

6.1 Implications

When planning experiments on multiple sections of the same course, researchers should be aware of the differences between sections that may have an influence on the student's grades independent of the instructional techniques used, and plan accordingly. If they must, for some reason or another, conduct an educational experiment between sections that are taught at different times of day, or under other differing circumstances, they should be aware of the typical variance in grades that can be brought on by such circumstances, and ensure that the effect size of the intervention they are trying to study is significantly larger. Alternatively, if one section

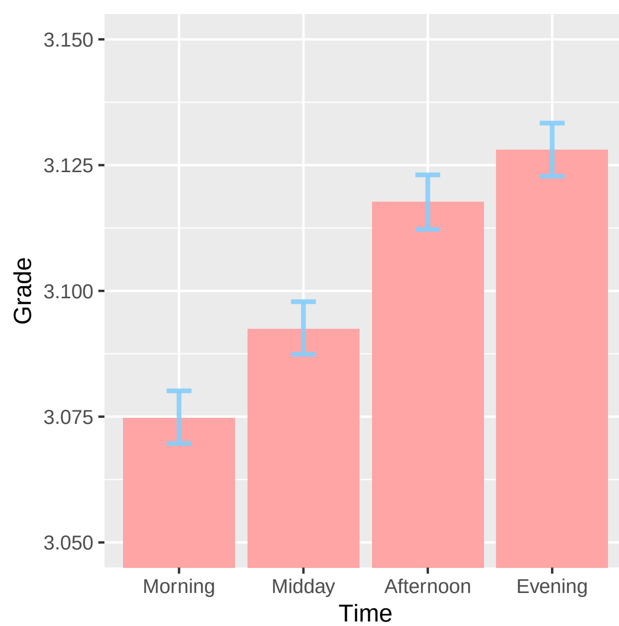

Figure 4: Simulated average grade if all courses were offered at the same time of day. Note that the y-axis does not start at 0. The 95% confidence intervals shown are basic bootstrap confidence intervals calculated using the distribution of student grades predicted by the model.

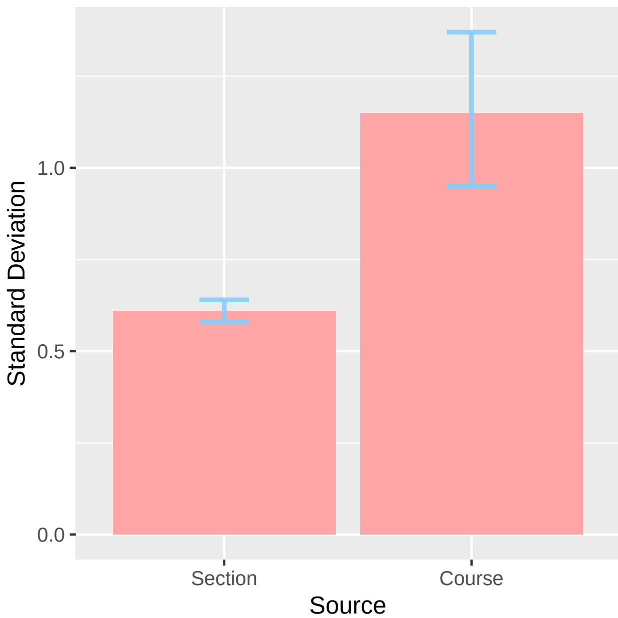

Figure 5: Standard deviation of the section level (U_{jk}) and course level (W_k) error terms from the model, with the 95% credible intervals provided by the Bayesian estimation.

is expected to have a higher grade due to documented reasons (i.e. being in the afternoon vs. in the morning), they could use the expected-to-be better performing section as the control, and the expected-to-be worse section as the experimental group, counting on the intervention to have a large enough effect to overcome the small negative impact of scheduling.

Despite what we do know about trends in student performance based on scheduling, it is critical to remember that all the above statistics only show general trends, and can not tell us about the relationship between any particular two course or section instances. Researchers should do all they can to ensure "all things equal" between their experimental and control groups, and should document all the information that they can about their course sections in the interest of good science, i.e. interpretability and reproducibility of their work. They should also be aware of and document the performance trends of the course they are experimenting with in particular, as some courses have much larger differences term-to-term or based on time of day than others do.

6.2 Limitations

As we have discussed, our study was limited by the data we were able to receive from the registrar at the University of Illinois at Urbana-Champaign, and the way that the computer science department decided to schedule the courses, making it impossible for us to draw any conclusions about day-of-week effects on scheduling. We also did not have access to attendance data, making it impossible to verify if morning classes performed more poorly for the same reason as in [17], namely, that students miss morning classes more often than they miss afternoon and evening classes.

Another limitation is that our data come from a single department at a single university. Replications of our study using data from other universities will be useful to corroborate our findings and give education researchers more confidence in the way they plan their experiments.

Additionally, our results should *not* be interpreted to mean that a particular student will earn higher grades if they register for afternoon classes instead of morning, because we are using observational data where students have self-selected into courses, leading to selection bias. Our study is unable to make any statement about *why* student performance varies by time of day, but a great area of future work would be to investigate why these performance differences exist, and what types of interventions may be able to help mitigate them.

7. CONCLUSION

We find that in the computer science department, the odds of a student receiving a higher grade in an afternoon class is 1.15 times the odds of a student with the same GPA in a morning class earning a higher grade. According to simulations run using our model, this difference amounts to an average grade difference of 0.04 grade points between morning and afternoon classes. There is also a large unexplained variance in grades between sections of the same course. Based on these findings and prior work in this area, we assert that the course scheduling information is an important piece of data which should be included in studies that make comparisons between treatments on different course sections. Based on our data set, we were not able to investigate trends in student performance based on which days of the week courses were scheduled for, and we found no overall trends for the term a course was offered in, or for the classroom building in which it was offered. Replication of our work, as well as work to answer the research questions which were unable to answer given our data set, would be great future contributions to the literature.

8. ACKNOWLEDGMENTS

We would like to thanks the computer science education research group at the University of Illinois at Urbana-Champaign for useful feedback and suggested references for this work.

9. REFERENCES

[1] A. M. Austin and L. Gustafson. Impact of course length on student learning. *Journal of Economics and Finance Education*, 5(1):26–37, 2006.

[2] K. Blaha, A. Monge, D. Sanders, B. Simon, and T. VanDeGrift. Do students recognize ambiguity in software design? a multi-national, multi-institutional report. In *Proceedings. 27th International Conference on Software Engineering, 2005. ICSE 2005.*, pages 615–616. IEEE, 2005.

[3] P.-C. BÃijrkner. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28, 2017.

[4] L. G. Carrington. The impact of course scheduling on student success in intermediate accounting. *American Journal of Business Education (AJBE)*, 3(4):51–60, 2010.

[5] M. A. Carskadon, C. Vieira, and C. Acebo. Association between puberty and delayed phase preference. *Sleep*, 16(3):258–262, 1993.

[6] T. Clear, J. Whalley, P. Robbins, A. Philpott, A. Eckerdal, and M.-J. Laakso. Report on the final bracelet workshop: Auckland university of technology, september 2010. 2011.

[7] J. Corley, A. Stanescu, L. Baumstark, and M. C. Orsega. Paper or ide? the impact of exam format on student performance in a cs1 course. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, SIGCSE âĂŹ20, page 706âĂŞ712, New York, NY, USA, 2020. Association for Computing Machinery.

[8] S. Fincher, B. Baker, I. Box, Q. Cutts, M. de Raadt, P. Haden, J. Hamer, M. Hamilton, R. Lister, and M. Petre. Programmed to succeed?: A multi-national, multi-institutional study of introductory programming courses. 2005.

[9] S. Fincher, R. Lister, T. Clear, A. Robins, J. Tenenberg, and M. Petre. Multi-institutional, multi-national studies in csed research: some design considerations and trade-offs. In *Proceedings of the first international workshop on Computing education research*, pages 111–121, 2005.

[10] S. Fincher, M. Petre, J. Tenenberg, K. Blaha, D. Bouvier, T.-Y. Chen, D. Chinn, S. Cooper, A. Eckerdal, and H. Johnson. A multi-national, multi-institutional study of student-generated software. *Kolin Kolistelut-Koli Calling Proceedings*, pages 20–28, 2004.

[11] M. H. Hagenauer, J. I. Perryman, T. M. Lee, and M. A. Carskadon. Adolescent changes in the homeostatic and circadian regulation of sleep. *Developmental neuroscience*, 31(4):276–284, 2009.

[12] G. L. Herman and S. Azad. A comparison of peer instruction and collaborative problem solving in a computer architecture course. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, SIGCSE âĂŹ20, page 461âĂŞ467, New York, NY, USA, 2020. Association for Computing Machinery.

[13] S. Krause-Levy, L. Porter, B. Simon, and C. Alvarado. Investigating the impact of employing multiple interventions in a cs1 course. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, SIGCSE âĂŹ20, page 1082âĂŞ1088, New York, NY, USA, 2020. Association for Computing Machinery.

[14] E. Kuo, M. M. Hull, A. Elby, and A. Gupta. Mathematical sensemaking as seeking coherence between calculations and concepts: Instruction and assessments for introductory physics. *arXiv preprint arXiv:1903.05596*, 2019.

[15] A. Lionelle, J. Grinslad, and J. R. Beveridge. Cs 0: Culture and coding. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, SIGCSE âĂŹ20, page 227âĂŞ233, New York, NY, USA, 2020. Association for Computing Machinery.

[16] R. Mansbach, A. Ferguson, K. Kilian, J. Krogstad, C. Leal, A. Schleife, D. Trinkle, M. West, and G. Herman. Reforming an undergraduate materials science curriculum with computational modules. *Journal of Materials Education*, 38(3-4):161–174, 2016.

[17] F. Marbouti, A. Shafaat, J. Ulas, and H. A. Diefes-Dux. Relationship between time of class and student grades in an active learning course. *Journal of Engineering Education*, 107(3):468–490, 2018.

[18] M. McCracken, V. Almstrum, D. Diaz, M. Guzdial, D. Hagan, Y. B.-D. Kolikant, C. Laxer, L. Thomas, I. Utting, and T. Wilusz. A multi-national, multi-institutional study of assessment of programming skills of first-year cs students. In *Working group reports from ITiCSE on Innovation and technology in computer science education*, pages 125–180. 2001.

[19] A. Naeem Syeda, R. Engineer, and B. Simion. Analyzing the effects of active learning classrooms in cs2. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, SIGCSE âĂŹ20, page 93âĂŞ99, New York, NY, USA, 2020. Association for Computing Machinery.

[20] J. Parham-Mocello and M. Erwig. Does story programming prepare for coding? In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, SIGCSE âĂŹ20, page 100âĂŞ106, New York, NY, USA, 2020. Association for Computing Machinery.

[21] M. S. Peteranetz and L.-K. Soh. A multi-level analysis of the relationship between instructional practices and retention in computer science. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, SIGCSE âĂŹ20, page 37âĂŞ43, New York, NY, USA, 2020. Association for Computing Machinery.

[22] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[23] K. Sanders, S. Fincher, D. Bouvier, G. Lewandowski, B. Morrison, L. Murphy, M. Petre, B. Richards, J. Tenenberg, and L. Thomas. A multi-institutional, multinational study of programming concepts using card sort data. *Expert Systems*, 22(3):121–128, 2005.

[24] T. A. Snijders and R. J. Bosker. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage, 2011.

[25] G. Sprint and E. Fox. Improving student study choices in cs1 with gamification and flipped classrooms. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, SIGCSE âĂŹ20, page 773âĂŞ779, New York, NY, USA, 2020. Association for Computing Machinery.

[26] M. Stieff, B. L. Dixon, M. Ryu, B. C. Kumi, and M. Hegarty. Strategy training eliminates sex differences in spatial problem solving in a stem domain. *Journal of Educational Psychology*, 106(2):390, 2014.