

# Studying Adaptive Learning Efficacy using Propensity Score Matching

Shirin Mojarad<sup>1</sup>, Alfred Essa<sup>1</sup>, Shahin Mojarad<sup>1</sup>, Ryan S. Baker<sup>2</sup>

McGraw-Hill Education<sup>1</sup>, University of Pennsylvania<sup>2</sup>

{shirin.mojarad, Alfred.essa, [s.a.mojarad](mailto:s.a.mojarad@mheducation.com)}@mheducation.com, Rybaker@upenn.com

**ABSTRACT:** Many higher education institutions are in the process of adopting adaptive learning platforms for online and hybrid learning. However, it is unclear how effective these platforms are at improving student success rates. ALEKS (Assessment and Learning in Knowledge Spaces) is an adaptive learning system designed for courses in science and mathematics. ALEKS has several mathematics courses that cover developmental mathematics for both four year and two year colleges. In this study, we investigate the effectiveness of ALEKS at a community college, where some sections have adopted ALEKS while others chose not to use it. We conduct different possible comparisons of ALEKS versus Non-ALEKS sections and students, including conducting a quasi-experiment using propensity score matching (PSM) to construct two similar groups of learners to compare between. PSM is conducted by matching ALEKS and Non-ALEKS users across their Accuplacer score, age and race. In all comparisons, students using ALEKS have significantly higher pass rates than comparison groups. When matching students using PSM, students who use ALEKS pass 19 percentage points more often than students who do not use ALEKS.

**Keywords:** efficacy, adaptive learning, ALEKS, propensity score matching, quasi-experimental design.

## 1 INTRODUCTION

Adaptive learning is increasingly being adopted across different institutions and disciplines to improve student outcome (Kolb & Kolb, 2005). However, it remains unclear how effective many of these platforms are at producing positive outcomes in these settings. It is important to investigate the efficacy of adaptive platforms in different educational settings, to help instructors, institutions, and students to decide which platforms to use in their classes (Hallahan, Keller, McKinney, Lloyd, & Bryan, 1988).

In this study, we examine the effectiveness of ALEKS (Assessment and Learning in Knowledge Spaces), a widely used adaptive learning system, in the context of a community college. Although there is evidence for ALEKS's effectiveness in other contexts, including in K-12 schools (Craig et al., 2013) and in non-traditional adult learning settings (Rivera, Davis, Feldman, & Rachkowski, 2017), there is not yet solid evidence on the efficacy of ALEKS in a community college setting.

We investigate this question within the context of a large community college in the Midwestern United States. According to many, a randomized controlled trial (RCT) would be the gold standard, ideal way to investigate this question (Silverman, 2009). However, RCTs are costly, and sometimes not feasible to conduct in educational settings, both due to difficulty in securing agreement for randomized assignment, and due to challenges in establishing implementation fidelity (Feng,

Roschelle, Heffernan, Fairman, & Murphy, 2014). For these reasons, many researchers have argued for the use of quasi-experiment studies besides or instead of RCTs. In these studies, subjects are assigned to treatment and control groups based on some criteria such as subjects' date of birth, while in RCTs this assignment is random. Quasi-experiment studies are a practical and acceptable alternative to RCTs when design of an RCT is implausible (Sullivan, 2011) .

Within this specific community college, it was not practical to randomly assign instructors or classes to conditions, as the college has made a policy decision that eliminates the ability to use an RCT design to study the efficacy of its chosen product. Instead, the college's administration decided that instructors would be given the choice of adopting ALEKS in their courses, and many instructors chose not to use it. Even when an instructor did choose to adopt ALEKS, it was not required for students, and was counted minimally towards the final grade. Therefore, only a portion of students in classes adopting ALEKS ever used ALEKS, and many students may not have used ALEKS to the degree or in the fashion intended. Therefore, we probably should not simply compare ALEKS classes to non-ALEKS classes; there are both selection bias issues and valid concerns about implementation fidelity (Feng et al., 2014).

An alternative would be to simply compare between the students who used and did not use ALEKS, ignoring what classes they were in. However, since this study was not designed as an RCT, there are issues of selection bias in making this comparison – it is possible, for example, that the students who decided to use ALEKS could have been the strong students to begin with and would have done well in the course anyways.

Therefore, in addition to investigating the effectiveness of ALEKS by comparing between different naturally-occurring student populations, we design a quasi-experiment study to isolate the effects of student characteristics and find comparable student populations who mainly differ in their use of ALEKS (P. R. Rosenbaum, 2010). This is done using propensity score matching (PSM) (Austin, 2011). Propensity score matching is explained in more detail in section 2. Section 3 explains the data used in this study and the study design. Sections 4 and 5 cover the results and conclusions of the study.

## **2 PROPENSITY SCORE MATCHING**

In RCT, random allocation is used to choose the treatment and control groups, so that study subjects have the same chance of being assigned to each study group. However, as described in section 1, random assignment is not plausible in many studies. For example, in the current study, learners are assigned to the ALEKS or non-ALEKS group depending on the class they have registered for and whether they chose to use ALEKS. The methods to study such groups are often described as quasi-experimental (Cochran & Chambers, 1965). The main concern in these studies is selection bias, since subjects are assigned to each group based on a criteria, and therefore they might not have similar chances of being assigned to treatment and control groups. PSM is a method that is used to remove this bias by finding control and treatment groups from the study cohort, such that they have similar probability of being assigned to the control and treatment group, at least according to a set of "baseline characteristic" variables describing members of the population. Therefore, it creates a study that resembles an RCT.

Considering two possible outcomes of receiving and not receiving treatment, each learner has two potential outcomes of  $Y_i(0)$  and  $Y_i(1)$ , the outcomes under the control and treatment, respectively. However, each learner is either in the control or treatment group. We define  $Z$  as an indicator variable on whether the learner received the treatment ( $Z = 0$  for control/non-ALEKS vs.  $Z = 1$  for treatment/ALEKS). Therefore, we can only observe one outcome for each learner.

In PSM, for each member of the intervention group, we identify a member of the control group that is as similar as possible in terms of their propensity score. Then, the difference in outcomes between the matched pair is computed. The average of this difference over the observed pairs is an estimate of the mean causal effect of a particular intervention on outcome.

A propensity score is used to choose treatment and control groups with similar baseline characteristics. A propensity score is defined as the probability of the subjects being assigned to the treatment group, given a set of baseline characteristics (P. Rosenbaum & Rubin, 1983). This can be formulated as the conditional probability of being exposed to intervention given baseline characteristics  $X$ :

$$e_i = P(Z_i = 1 | X_i)$$

where  $e_i$  is the propensity score and  $X$  is the vector of observed characteristics of the subject. This can be modelled as a logistic regression model, where the dependent variable is the probability of receiving treatment and independent variables are the baseline characteristics:

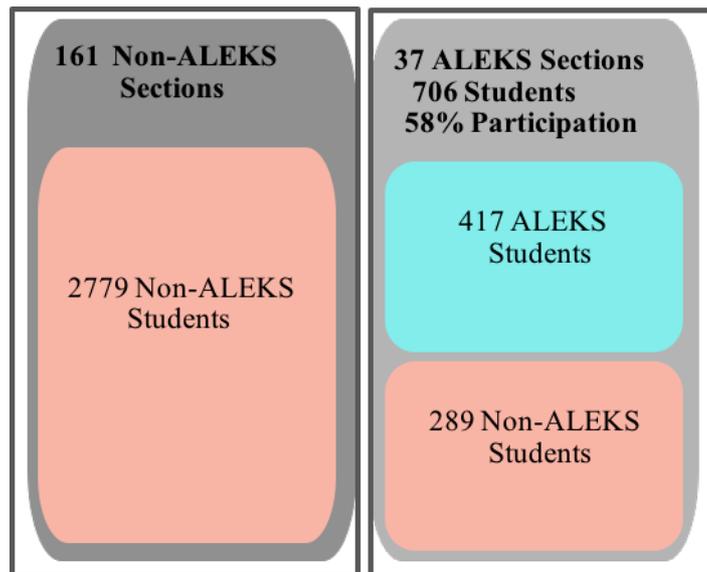
$$p(Z = 1 | X) = \frac{1}{1 + e^{-\beta X}}$$

One advantage of PSM is that the regression model used to predict the probability of receiving treatment takes into account the relationship between baseline characteristics. In addition, PSM enables matching not just at the mean but balances the distribution of observed characteristics across treatment and control.

### 3 DATA AND METHODOLOGY

#### 3.1 Data

We obtained data from 3422 students in 198 sections covering four courses including pre-algebra (67 sections), elementary algebra (44 sections), intermediate algebra (43 sections) and college math (44 sections). Amongst these, 37 sections with 706 students adopted ALEKS. From these students, only 417 (59%) used ALEKS. Figure 1 shows a representation of ALEKS and non-ALEKS sections and students.



**Figure 1: Breakdown of ALEKS and non-ALEKS sections and students.**

### 3.2 Methodology

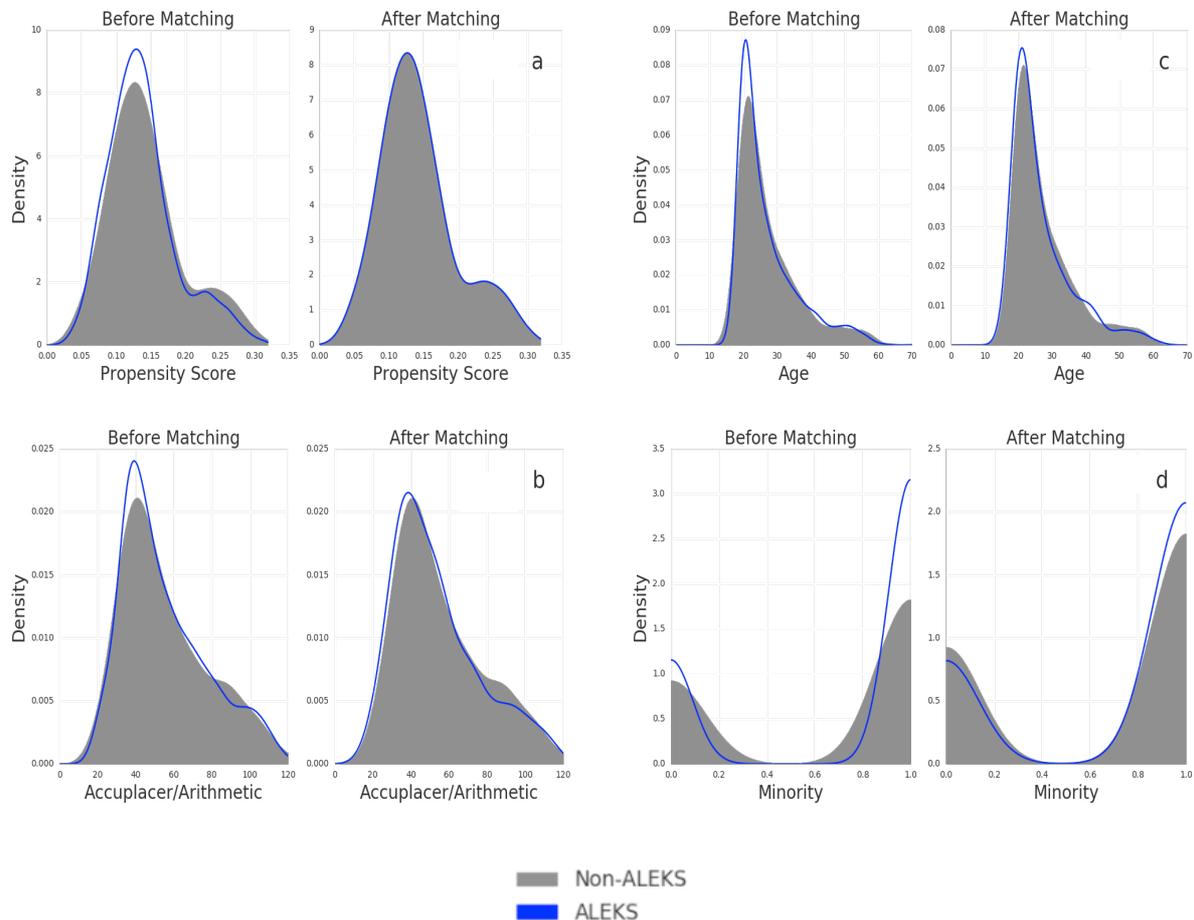
We have made comparisons of several possible breakdowns of ALEKS vs. Non-ALEKS students and sections. Below are the comparisons conducted in this study:

1. ALEKS students vs. all Non-ALEKS students (in both ALEKS and non-ALEKS sections)
2. ALEKS students in ALEKS sections vs. Non-ALEKS students in ALEKS sections
3. ALEKS students in ALEKS sections vs. Non-ALEKS students in Non-ALEKS sections
4. ALEKS sections vs. Non-ALEKS sections
5. Matched ALEKS students vs. Matched Non-ALEKS students

What could differentiate students in comparison groups 1-4 is their starting knowledge and their current learning situation which could be affected by their age and race. Therefore, the last comparison is done by matching ALEKS and non-ALEKS users using PSM. Matching is done using three student characteristics: Accuplacer arithmetic score, age, and whether the student's race is classified as minority or not. The Accuplacer score is used by the college to decide whether to place students into developmental math courses and is used as a measure of students' prior knowledge in the subject (Mattern & Packman, 2009). The student group from which the control matches were identified includes only non-ALEKS students in non-ALEKS sections. This naturally removes the student selection bias in the control group, since students in non-ALEKS sections do not have a choice to use ALEKS.

A logistic regression model is used to calculate the propensity score of students -- specifically, the binomial generalized linear model from *statsmodels* package in Python was used. The logistic regression model had whether the student used ALEKS as a binary outcome and the independent attributes consisted of Accuplacer arithmetic score, age, and whether the student race is minority or

not. Figures 2.a-2.d shows the distribution of each of these attributes and the propensity score of ALEKS and non-ALEKS users before and after matching. Matching on propensity score is conducted as a 1-1 matching using nearest neighbor approach, which uses the distance between propensity scores to find the closest match. Hence, for each treatment subject, a control match is selected as the subjects with the closest propensity score.



**Figure 2: distribution of a) propensity score b) Accuplacer score c) age, d) minority, before and after matching for ALEKS (blue) and non-ALEKS (grey) users.**

## 4 RESULTS

Table 1 shows the ALEKS and non-ALEKS group pass rates, the ALEKS-non-ALEKS group difference in pass rates, and p-value for each of the comparisons mentioned in section 3.2. The criteria for pass is grades C+ and above. Students' grades are measured using a uniform test conducted across all classes at the end of semester. We have used the chi-square ( $\chi^2$ ) contingency test to compare the pass rate across two groups (Rao & Scott, n.d.). We conduct five comparisons. The first comparison is all students who at least took an initial assessment in ALEKS (ALEKS students) versus all students who did not use ALEKS in the course of the class (non-ALEKS students). Within this comparison, ALEKS students had statistically significantly higher pass rates,  $\chi^2(df=1, N=3422) = 29.5$ ,  $p < 0.001$ , with ALEKS achieving a boost of 14 points in pass rates.

We have used the chi-square ( $\chi^2$ ) contingency test to compare the pass rate across two groups (Rao & Scott, n.d.). We conduct five comparisons. The first comparison is all students who at least took an initial assessment in ALEKS (ALEKS students) versus all students who did not use ALEKS in the course of the class (non-ALEKS students). Within this comparison, ALEKS students had statistically significantly higher pass rates,  $\chi^2(df=1, N=3422) = 29.5, p<0.001$ , with ALEKS achieving a boost of 14 points in pass rates.

The second comparison is between ALEKS and non-ALEKS only within ALEKS sections. This comparison is important as it naturally controls for the instructor and class environment, by comparing students who did and did not use ALEKS within the same class. Within this comparison, ALEKS students had statistically significantly higher pass rates,  $\chi^2(df=1, N=709) = 24.7, p<0.001$ , with ALEKS achieving a boost of 19 points in pass rates.

The third comparison considers assignment at the classroom level. In this comparison, all students within ALEKS sections, whether they did or did not use ALEKS, are compared against all the students in non-ALEKS sections. This comparison is perhaps the most standard quasi-experimental comparison, but raises questions of implementation fidelity. Within this comparison, ALEKS students had statistically significantly higher pass rates,  $\chi^2(df=1, N=198) = 8.1, p=0.004$ , with ALEKS achieving a boost of 6 points in pass rates.

The fourth comparison is between ALEKS students in ALEKS sections and non-ALEKS students in non-ALEKS sections. Within this comparison, we are excluding non-ALEKS students in ALEKS sections from this comparison as those are the students who chose not to use ALEKS, despite having the option of using it in the class. Including these students includes students who did not participate in the treatment, despite being assigned to the treatment group, creating questions of implementation fidelity. Within this comparison, ALEKS students had statistically significantly higher pass rates,  $\chi^2(df=1, N=3196) = 27.5, p<0.001$ , with ALEKS achieving a boost of 14 points in pass rates.

Finally, comparison five attempts to avoid the biases inherent in the first four comparisons, comparing ALEKS students who are matched with similar non-ALEKS students in non-ALEKS classes. The matching is done using Accuplacer, age and minority and as shown above, the students selected in the matching process have similar prior knowledge, age, and minority between conditions. All students in the matched treatment condition used ALEKS and all students in the matched control condition did not use ALEKS. Within this comparison, ALEKS students had statistically significantly higher pass rates,  $\chi^2(df=1, N=748) = 16.5, p<0.001$ , with ALEKS achieving a boost of 15 points in pass rates.

As shown in this table, all comparisons are statistically significantly in favor of ALEKS, with a boost of 6 to 19 points in pass rates between ALEKS and non-ALEKS users across different comparisons. Some of the comparisons are likely to be biased in favor of ALEKS, others against ALEKS, but overall they tell a common story – ALEKS is statistically significantly more effective at enhancing pass rates compared to the control condition.

**Table 1: Pass rates and significance level for ALEKS and non-ALEKS users.**

Comparison	Pass Rates		
	for ALEKS vs. Non-ALEKS	Boost	p-value
1. ALEKS students vs. all Non-ALEKS students	71% vs 57%	+14	<0.001
2. ALEKS students vs. Non-ALEKS students in ALEKS sections	71% vs 52%	+19	<0.001
3. ALEKS sections vs. Non-ALEKS sections	63% vs 57%	+6	0.004
4. ALEKS students in ALEKS sections vs. Non-ALEKS students in Non-ALEKS sections	71% vs 57%	+14	<0.001
5. Matched ALEKS students vs. Matched Non-ALEKS students (quasi-experiment study using Propensity Score Matching)	70% vs 55%	+15	<0.001

## 5 CONCLUSION

In this paper, we present a study on the pass/fail outcomes of students who used and did not use ALEKS within their developmental math courses. For several comparisons, the results show significantly higher pass rates amongst students using ALEKS. However, it may not be valid to compare between groups directly, due to concerns around selection bias and implementation fidelity. Therefore, we conducted a quasi-experiment study using Propensity Score Matching (PSM), labeled comparison 5 in Table 1. The results show that students using ALEKS have significantly higher pass rates, even when we use PSM to control for students' math placement score, age and race. However, as with all PSM-based quasi-experiment study, it may be that the matching was imperfect. It is important to consider other factors that could affect student outcome besides the ones used for matching ALEKS and non-ALEKS users in this study. Other factors could include but are not limited to social-economic background, and prior academic performance, as well as students' attitudes towards learning and attitudes towards online learning technologies. As such, conducting further follow-up studies will help us more conclusively understand whether ALEKS is positively benefitting students.

Following the results of this study, the college in which we conducted the study is in process of adopting ALEKS in more sections, and encouraging instructors to make ALEKS a requirement for students and a part of their grades. We intend to follow-up this study with a subsequent study, at the same institution, to see if usage has genuinely increased, and if so, whether the greater proportion of ALEKS users maintain the same improvements in outcomes seen in this study. Within this upcoming study, we intend to also control for a greater range of factors. By doing so, we may be able to better understand the degree to which ALEKS is benefitting students, and whether these benefits are equivalent across all groups of students.

## ACKNOWLEDGMENTS

[Redacted for submission]

## REFERENCES

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of

- Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>
- Cochran, W. G., & Chambers, S. P. (1965). The Planning of Observational Studies of Human Populations. *Journal of the Royal Statistical Society. Series A (General)*, 128(2), 234. <https://doi.org/10.2307/2344179>
- Craig, S. D., Hu, X., Graesser, A. C., Bargagliotti, A. E., Sterbinsky, A., Cheney, K. R., & Okwumabua, T. (2013). The impact of a technology-based mathematics after-school program using ALEKS on student's knowledge and behaviors. *Computers and Education*, 68(October), 495–504. <https://doi.org/10.1016/j.compedu.2013.06.010>
- Feng, M., Roschelle, J., Heffernan, N., Fairman, J., & Murphy, R. (2014). Implementation of an Intelligent Tutoring System for Online Homework Support in an Efficacy Trial (pp. 561–566). Springer, Cham. [https://doi.org/10.1007/978-3-319-07221-0\\_71](https://doi.org/10.1007/978-3-319-07221-0_71)
- Hallahan, D. P., Keller, C. E., McKinney, J. D., Lloyd, J. W., & Bryan, T. (1988). Examining the Research Base of the Regular Education Initiative. *Journal of Learning Disabilities*, 21(1), 29–35. <https://doi.org/10.1177/002221948802100106>
- Kolb, A. Y., & Kolb, D. A. (2005). Learning Styles and Learning Spaces: Enhancing Experiential Learning in Higher Education. *Academy of Management Learning & Education*, 4(2), 193–212. <https://doi.org/10.5465/AMLE.2005.17268566>
- Mattern, K. D., & Packman, S. (2009, December 4). Predictive Validity of ACCUPLACER Scores for Course Placement: A Meta-Analysis. Research report, The College Board. Retrieved from <https://research.collegeboard.org/publications/content/2012/05/predictive-validity-accuplacer-scores-course-placement-meta-analysis>
- Rao, J. N. K., & Scott, A. J. (n.d.). On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. *The Annals of Statistics*. Institute of Mathematical Statistics. <https://doi.org/10.2307/2241033>
- Rivera, M. A., Davis, M. H., Feldman, A., & Rachkowski, C. (2017). An outcome evaluation of an adult education and postsecondary alignment program: the Accelerate New Mexico experience. *Problems and Perspectives in Management (Open-Access)*, 11(4). Retrieved from <https://businessperspectives.org/journals/problems-and-perspectives-in-management/issue-40/an-outcome-evaluation-of-an-adult-education-and-postsecondary-alignment-program-the-accelerate-new-mexico-experience>
- Rosenbaum, P. R. (2010). *Design of Observational Studies*. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4419-1213-8>
- Rosenbaum, P., & Rubin, D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70, 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Sullivan, G. M. (2011). Getting Off the “Gold Standard”: Randomized Controlled Trials and Education Research. *Journal of Graduate Medical Education*, 3(3), 285–289. <https://doi.org/10.4300/JGME-D-11-00147.1>