Can a Single Placement Test Effectively Place Students Into Different Math Pathways?

Kristin Frank
Towson University

Alexei Kolesnikov
Towson University

Xiaoyin Wang
Towson University

Math placement policies are intended to match incoming freshmen with an appropriate math course. At our four-year institution, the same placement test is used to place students into nine courses across three pathways. In this paper we use institutional data to analyze the effectiveness of test-based placement into different courses requiring similar (and sometimes the same) cutscore. We use the classification power metric to evaluate and compare the effectiveness of three placement policies. We report that that placement tests provide minimal information about students' readiness for mathematics courses that are not algebra intensive, such as basic statistics.

Keywords: Placement, Classification Power, Math Pathways

Improving student success in entry-level college mathematics courses is a high priority for many higher education institutions. One of the approaches to improving student success involves developing and refining the process that matches incoming students with an appropriate mathematics course. In many cases, this process involves incoming students taking a placement test where the students' performance on the test, possibly in conjunction with other measures like high school GPA, determines which mathematics course(s) they are eligible to take.

Much of the research on mathematics placement has been conducted in the context of two-year institutions with a focus on the distinction between remedial and college-level coursework. Our study extends this body of research to evaluate placement at a four-year institution utilizing a test-based multi-level placement system. Placement policies at our institution are intended to simultaneously assess a student's readiness for college-level mathematics and differentiate between a student's readiness for various levels of mathematics courses. To further complicate the placement decision, there are three mathematics pathways students can choose depending on their major and career goals. Thus, the same placement test score can place students in different college-level courses. In this paper we use the classification power metric to compare the effectiveness of multi-level placement policies based on different placement-tests. Our results highlight that the placement process at our institution provides little information about students' readiness for non-algebra intensive courses.

Background

There are two main strategies to evaluate the effectiveness of a placement policy: comparative analysis of observable outcomes and classification error analysis. Researchers conducting comparative analysis of observable outcomes focus on the relationship between placement decisions and observable outcomes such as students' grades in their first math course, retention, and time to graduation (see, for example, Barnett et al., 2018, Chen, 2016, Rodriguez, 2014; and Thornton et al., 2019). Frank et al. (2022) found that the placement decision has low explanatory power for both short- and long-term outcome variables and argued that such comparative analysis of observable outcomes is not sufficient to meaningfully evaluate placement policies.

Another strategy is to treat the placement decision as a classification problem where the goal of placement is to identify, with as high accuracy as possible, the students who are ready to take the target course and those who need to take a prerequisite (see Belfied & Crosta, 2012; Scott-Clayton, 2012). The quality of the decision can be measured with the *severe error rate* (SER). Scott-Clayton (2012) defines the SER as the sum of the percentage of students who were placed into developmental math but would have been able to pass college-level math with the grade of B or higher and the percentage of students who were placed into college-level math but earned a failing grade (F). Researchers report SERs between 16% and 28% in the context of mathematics placement at two-year institutions (Belfield & Crosta, 2012; Scott-Clayton, 2012).

Frank et al. (in press) identified three methodological challenges associated with using the SER to evaluate and compare placement decisions: (1) challenges associated with determining whether a given SER is evidence of an effective placement policy when there is no benchmark for comparison, (2) challenges associated with comparing SERs across placement policies; and (3) interpretational challenges due to possible variation in the magnitude of the SER due to chance. They propose a new metric, the *classification power* (CP), which quantifies the amount of information provided by the placement policy about students' readiness for a given course. The CP measures the reduction in the number of severe errors from implementing a given policy instead of a hypothetical placement policy that provides no information about students' readiness (imagined as a weighted coin-flip). Since the classification power compares the number of severe errors between the implemented policy and a hypothetical policy that places the same cohort of students, the classification power can be compared across courses and across placement periods.

Since the classification power measures the reduction in severe errors, the greater the classification power the more effective the policy. A policy is deemed ineffective if the 95% confidence interval for its classification power includes 0 since a classification power of 0 indicates that the implemented policy provides no improvement compared to a hypothetical policy that can be imagined as using a weighted coin-flip to place students.

To inform policy decisions, it is advantageous to consider separately the CP for both ready students (i.e., students who are able to earn a grade of B or better in a given course without taking the prerequisite) and not ready students (i.e., students would earn a failing grade in a given course if they do not first take a prerequisite). The CP for ready students (CP_R) can be interpreted as the percentage of ready students appropriately placed by the placement policy while the remaining ready students are imagined to be placed with a weighted coin-flip where the coin is weighted to match the overall placement rate. Similarly, the CP for not ready students (CP_{NR}) can be interpreted as the percentage of not ready students appropriately placed into a prerequisite while the remaining not ready students are imagined to be placed with a weighted coin-flip. Frank et al. (2024) analyze placement into Precalculus at their 4-year institution. They report lower classification powers for not ready students suggesting that placement at their institution is better at identifying ready students than not ready students.

Methodology

To evaluate and compare placement policies that place students into multiple college-level courses (e.g., Basic Statistics, College Algebra, Precalculus, etc.), we determine the CP for a single course at a time. We call this the *target course*. Each student who eventually takes the target course as part of their program of study – the *target course population* – was either placed or not placed into the target course when matriculating into our four-year institution. A student is placed into the target course when they have met the qualifying score on the placement test. Our data consists of institutional records for students in the target course population including measures of academic preparation (high school GPA, SAT/ACT) as well as demographic factors (race, gender, age).

In order to estimate the unobserved severe errors, we follow the process described by Scott-Clayton (2012) to perform a statistical matching between students who took the target course without taking prerequisites and those who took the target course after a prerequisite. The process is fully described in (Frank et al., in press). We fit two regression models on the group of students who took the course without prerequisites. The first model estimates the probability of a student earning higher than a B in their first attempt in the target course given a vector of academic and demographic data. The model is applied to students who earned at least a B in the target course on their first attempt. Students with a predicted probability of passing over 50% are classified as underplaced; this produces an estimate of unobserved underplaced students. The second model estimates the probability a student fails the target course without prerequisites; this model is applied to students who took a prerequisite and earned at least a D on their first attempt in the target course. Students with a predicted probability of failing over 50% are classified as overplaced. These estimates, together with the students observed to be severely misplaced give an estimate for the SER since the SER is the sum of all observed and estimated severe errors.

The classification power is defined as the reduction in the number of severe errors made by the implemented policy and a hypothetical policy where the placement decision is independent of a student's readiness (Frank et al., 2024). We use ρ_R (and ρ_{NR}) to denote the SER for ready (and not ready) students for this hypothetical policy. We separately compute CP_R and CP_{NR} where:

$$CP_R = 1 - \frac{\mathbf{P}(\text{Ready and Not Placed})}{\mathbf{P}(\text{Ready})\mathbf{P}(\text{Not Placed})} = 1 - \frac{SER_R}{\rho_R} \text{ and } CP_{NR} = 1 - \frac{\mathbf{P}(\text{Not Ready and Placed})}{\mathbf{P}(\text{Not Ready})\mathbf{P}(\text{Placed})} = 1 - \frac{SER_{NR}}{\rho_{NR}}.$$

The overall CP is a weighted average of CP_R and CP_{NR} where each metric is weighted by the relative proportion of severe errors made by the hypothetical policy.

Institutional Context and Data

At our four-year institution, a student's score on a math placement test and their declared major are used to place students into one of nine math courses, eight of which offer college credit. The courses are organized into three pathways: a non-algebra pathway culminating in basic statistics, an algebra pathway culminating in Calculus for Applications (a business calc course), and an algebra intensive pathway culminating in Calculus I (this is the only pathway requiring precalculus).

Since 2014, our institution has implemented two placement tests. Between 2014 and 2018 we administered the Mathematical Association of America's Basic Algebra Test (BAT) and Calculus Readiness Test (CRT). Students wishing to take either business calculus or calculus needed a qualifying score on the BAT (17) and a qualifying score on the CRT (see Table 1). In 2018 our institution adopted the Assessment and Learning in Knowledge Spaces Placement, Preparation, and Learning Test (ALEKS PPL[1]). Both placement tests were administered to all incoming students without college math credit in an online and unproctored environment. Cutscores for a subset of courses considered in this paper are provided in Table 1. Notice that in 2016 our institution changed the cutscores for the BAT/CRT test thus we consider separately the students placed under each set of cutscores.

[1] The ALEKS placement test is an adaptive test consisting of 25 questions designed to assess students' knowledge across 314 topics (McGraw Hill, 2024).

Table 1. Cutscores for BAT/CRT placement test and ALEKS placement test.

	BAT/CRT v1 F14-S16	BAT/CRT v2 F16-S18	ALEKS F18- present
Basic Statistics	BAT 17	BAT 14	ALEKS 30
College Algebra	BAT 12	BAT 14	ALEKS 46
Precalculus	BAT 17	BAT 20	ALEKS 61
Business Calculus	CRT 11	CRT 11	ALEKS 61
Calculus I	CRT 11	CRT 15	ALEKS 76

Recognizing that there are three math pathways at our institution, the same placement test score could place students into three different courses depending on their declared major. For example, an ALEKS score of 62 would place a sociology major into basic statistics, an economics major into business calculus, and a physics major into precalculus.

Data

Our data consists of the academic records for all first-time freshmen who enrolled in Towson University, a large public university in the mid-Atlantic region, between Fall 2014 and Fall 2018. We exclude students who enrolled in/after Fall 2019 due to confounding effects of the COVID pandemic. We evaluated the CP for four courses across three pathways (college algebra, precalculus, basic statistics, and business calculus). Table 2 summarizes the demographic characteristics of these target course populations and provides the percentage of students in the target course population who were placed into the target course by the placement test. We partition the data by cohort based on the placement policy at the time of a student's enrollment.

Table 2. Summary of demographic and academic preparation for each target course.

		Sample Size n	White %	Black %	Asian %	Hispanic %	Other %	Placement Rate %
College Algebra	BAT/ CRTv1	1,076	58.27	19.42	6.41	7.90	7.99	88.85
	BAT/ CRTv2	1,566	50.13	24.58	6.90	10.09	8.30	83.21
	ALEKS	850	48.35	29.18	8.24	7.06	7.18	74.71
Precalc	BAT/ CRTv1	679	54.64	21.94	8.25	7.66	7.51	82.62
	BAT/ CRTv2	483	38.51	33.54	9.32	9.11	9.52	60.46
	ALEKS	262	38.17	37.79	8.02	8.02	8.02	59.92
Basic Stats	BAT/ CRTv1	875	59.54	18.63	7.89	6.86	7.09	71.89
	BAT/ CRTv2	939	52.82	22.15	9.16	8.09	7.77	85.20
	ALEKS	582	45.53	29.38	9.45	8.25	7.39	95.02
Business Calc	BAT/ CRTv1	635	63.62	14.33	6.93	6.61	8.50	36.85
	BAT/ CRTv2	617	56.08	21.23	7.29	9.24	6.16	45.71
	ALEKS	336	51.79	25.60	7.14	6.55	8.93	60.71

We recognize some demographic shifts in the sample over time. We do not see this as problematic since we perform the analysis on each cohort of students and the demographic characteristics are more uniform within a cohort. Additionally, we note sample sizes greater than 260, sufficiently large to perform the statistical matching described in the methodology.

Results

In this section we provide the results of the statistical matching procedure described in the methodology. Tables 3 and 4 present: the CP, the CP for ready students (CP_R), and CP for not ready students (CP_{NR}) for four courses across math pathways. We also provide 95% confidence intervals (CI) computed using bootstrapping techniques. We selected these four courses from the eight credit-bearing courses because they share similar (sometimes identical) cutscores (see Table 1). Table 3 provides the results for two courses in the algebra-intensive pathway and Table 4 provides results for two courses not in the algebra-intensive pathway.

Table 3. Comparison of classification power for courses in the algebra intensive pathway.

		BAT/CRTv1		BAT/CRTv2		ALEKS	
		Est (%)	CI (%)	Est (%)	CI (%)	Est (%)	CI (%)
College Algebra	SER	7.2	(5.8,8.8)	10.2	(8.8,11.8)	12.8	(10.6,15.3)
	CP	34.7	(22.9,45.3)	25.9	(16.9,33.7)	30.7	(20.3,40.5)
	CP_R	40.3	(26.7,53.2)	29.1	(18.2,38.8)	39.4	(26.5,52.1)
	CP_{NR}	23.3	(8.3,38.6)	19.9	(8.4,31.0)	18.8	(5.8,31.7)
Precalc	SER	7.8	(5.9,10.0)	16.1	(12.6,20.1)	15.3	(10.7,20.6)
	CP	46.1	(33.6,57.5)	30.8	(18.0,43.2)	13.9	(-5.5,33.5)
	CP_R	59.7	(44.2,74.1)	33.6	(20.0,48.5)	24.3	(3.6,45.2)
	CP_{NR}	27.4	(11.7,43.3)	24.5	(3.7,43.7)	-20.5	(-54.3,17.1)

Table 4. Comparison of classification power for courses not in the algebra intensive pathway.

		BAT/CRTv1		BAT/CRTv2		ALEKS	
		Est (%)	CI (%)	Est (%)	CI (%)	Est (%)	CI (%)
Basic Stats	SER	17.3	(14.2,21.0)	10.6	(8.7,13.3)	12.9	(10.3,16.7)
	CP	-15.7	(-27.4,-1.4)	16.7	(5.4,27.7)	3.2	(-4.7,10.7)
	CP_R	-33.4	(-50.3,-8.0)	22.6	(-0.8,49.1)	19.1	(-49.8,81.8)
	CP_{NR}	0.8	(-13.8,13.4)	14.1	(3.4,23.4)	1.8	(-3.3,7.3)
Business Calc	SER	18.0	(14.0,22.0)	18.6	(15.4,23.2)	30.4	(24.7,35.7)
	CP	21.5	(10.6,32.8)	21.3	(8.2,32.0)	-13.3	(-25.7,1.1)
	CP_R	15.5	(6.2,27.7)	19.6	(7.0,34.0)	3.2	(-13.7,24.4)
	CP_{NR}	36.3	(11.7,54.8)	23.2	(2.6,38.9)	-31.8	(-48.7,-15.5)

Effectiveness of Test-Based Placement

Across these four courses and three placement policies we report estimates for the CP between -15.7 and 46.1. As we will discuss in the next section, we take this wide range in CPs across courses as evidence that the amount of information provided by a placement test varies based on the nature of the target course and the placement test administered.

Before comparing the effectiveness of a placement test across courses, we begin with a brief discussion of the evidence of test-based placement focusing our discussion on ALEKS based placement into precalculus. For ALEKS based placement into precalculus, we report a CP of 13.9 and associated CI of (-5.5, 33.5). We interpret this CI as evidence that we can expect substantial variability in the classification power due to chance and the specific characteristics of the enrolled student population. We note that we see large confidence intervals for the SER as well. We take this as evidence that small, single percentage point, variations in any measure of the effectiveness of a placement policy may be simply attributed to chance.

We see that for placement into precalculus both BAT/CRTv2 and ALEKS made about the same proportion of severe errors (SERs around 15%). However, we see different CPs for these policies; we report a CP of 30.8 for BAT/CRTv2 and 13.9 for ALEKS. This provides evidence that two policies can make the same proportion of severe errors while providing different information about students' readiness for a course. We attribute the difference in these estimates to variation in the percentage of students ready and not ready for the course in each cohort: the SER does not take into consideration differences in the sample characteristics while the CP attends to these differences by comparing an implemented policy to a hypothetical policy that places the same population of students.

We now consider the CP for ready and not ready students separately. For ALEKS based placement into precalculus we report an estimate for CP_R of 24.3 and an estimate for CP_{NR} of -20.5. The magnitude of these results can be interpreted as the percentage of ready (and not ready) students we confidently placed correctly. That is, we estimate that ALEKS based placement correctly placed 24.3% of ready students into precalculus and *incorrectly* placed 20.5% of students not ready for precalculus into the course while placing the remaining with a coin-flip weighted to match the placement rate (here 59.92%). That is, the negative estimate (for either the CP_R or CP_{NR}) implies the implemented policy made more severe errors than using the hypothetical coin-flip policy for all students.

We emphasize that the benchmark here is a weighted coin-flip policy – an obviously suboptimal placement policy. Looking across courses most often we report CPs for ready and not ready students around 25%. That is, most often we are confidently placing 25% of ready and not ready students. We take these reported CPs as minimal evidence for the effectiveness of test-based mathematics placement in providing information about students' readiness for a target course.

Comparison of Classification Power across Courses

At our institution, the same placement test is used to place students into nine different courses. In this section we evaluate the ability of a placement test, even an adaptive test like ALEKS PPL, to provide information about students' readiness for such a variety of courses. Specifically, we compare the effectiveness of placement into College Algebra and Basic Statistics as well as Precalculus and Business Calculus since these courses share similar (sometimes identical) cutscores across placement cohorts.

Across all three placement periods, the CP for Basic Statistics is less than the CP for College Algebra. These are significant differences for both BAT/CRTv1 and ALEKS as there is no overlap in the confidence intervals. Additionally, we report negative estimates for the CP for Basic Statistics under BAT/CRTv1 and ALEKS suggesting that a hypothetical policy using a weighted coin-flip would have made fewer errors than these implemented test-based policies. We also note that Basic Statistics has the lowest estimate for CP_R across all courses. Furthermore, this estimate is negative (-33.6). Together, we interpret this as preliminary evidence that test-based placement could serve as a gatekeeper in non-algebra intensive courses as the tests are incorrectly underplacing a larger percentage of ready students than a weighted coin-flip. We do not interpret this as evidence that there should be no placement policy for such courses as we estimate around 10% of students are not ready for basic statistics and would benefit from taking a prerequisite. Our results highlight the need for better placement mechanisms.

Next, we directly compare placement into precalculus and business calculus[2]. Across all three placement periods the CP for business calculus is less than the CP for precalculus. This difference is significant for placement under BAT/CRTv1; there is no statistically significant difference under the ALEKS policy which uses the same adaptive test and cutscore for both courses. This is noteworthy since under BAT/CRTv2 students took different placement tests to place into these courses, yet these different tests did provide statistically comparable amounts of information for each course. It is also worth noting that the CRT underplaced many students ready for business calculus as evidenced by the low CP_R (15.5). This suggests that utilizing different placement tests for different target courses is not necessarily an effective way to implement a multi-level placement policy.

Impact of Cutscore on Policy Effectiveness

We notice that changing the cutscores between BAT/CRTv1 and BAT/CRTv2 had different impacts across courses. For both college algebra and precalculus, increasing the cutscore reduced the percentage of students placed into the course (see Table 2). However, this change did not improve the policy's ability to identify students not ready for the course as we see no noticeable change in CP_{NR}: the slight decrease in the reported point estimates can be attributed to variation due to chance). In contrast, decreasing the cutscore for basic statistics did improve the policy's ability to identify ready students as we see an increase in CP_R. This is a noteworthy change since the initial cutscore of 17 on BATv1 resulted in a policy less effective than a weighted coin flip, as indicated by the negative estimate of CP_R. However, after decreasing the cutscore the estimates for CP_R and overall CP increased so that we can say the policy is more effective than the hypothetical coin-flip.

These results echo Leeds and Mokher's (2020) finding that changing the cutscore can impact placement accuracy. In fact, our data suggests that increasing the cutscore does not necessarily result in improved placement accuracy but can instead serve as a gatekeeper for ready students.

Discussion

At our four-year institution, a placement test is used to place students into nine courses across three different math pathways. We questioned the effectiveness of a single test (even an adaptive test) in placing students into multiple courses. Our findings provide preliminary evidence that test-based placement serves as a gatekeeper for courses outside the algebra intensive pathway (i.e., courses not leading to Calculus I). We do not interpret this as evidence that there should be no placement policy as we estimate that there are students not ready for these courses (about 10% of students who need to take basic statistics need a prerequisite and about 20% of students who need to take business calculus need a prerequisite).

We recognize that these findings could be the result of the established cutscores. Thus, we plan to use the institutional data to construct a proposed policy that sets cutscores to maximize the CP. Evaluating the CPs associated with such a proposed policy would allow us to distinguish whether it is the placement test or the cutscore that limits the information the test score is able to provide about students' readiness for a given course.

References

Barnett, E., Bergman, P., Kopko, E. M., Reddy, V. T., Belfield, C., & Roy, S. (2018). Multiple

[2] At our institution college algebra, not precalculus, is a prerequisite for business calculus. Thus, we consider business calculus to be on a less algebra-intensive pathway.

measures placement using data analytics: An implementation and early impacts report.
Chen, X. (2016). Remedial coursetaking at US public 2-and 4-year institutions: Scope, experiences, and outcomes. Statistical analysis report. NCES 2016-405. National Center for Education Statistics.
Belfield, C. R., & Crosta, P. M. (2012). Predicting success in college: The importance of placement tests and high school transcripts. CCRC working paper no. 42. Community College Research Center, Columbia University.
Frank, K., Kolesnikov, A., Wang, X. (2022). Comparing placement policies at a four-year institution. *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies, 32*(10). https://doi.org/10.1080/10511970.2021.1983903
Frank, K., Kolesnikov, A., Wang, X. (in press). The classification power: A metric for evaluating and comparing postsecondary mathematics placement policies. *Journal of Postsecondary Student Success.*
Leeds, D. M., & Mokher, C. G. (2020). Improving indicators of college readiness: Methods for optimally placing students into multiple levels of postsecondary coursework. *Educational Evaluation and Policy Analysi*s, 42(1), 87–109. https://doi.org/10.3102/0162373719885648
McGraw Hill (2024). *ALEKS PPL: Placement, Preparation and Learning*. ALEKS Placement, Preparation and Learning. https://www.mheducation.com/highered/alekspp l.html#accurate
Rodríguez, O. (2014). Increasing access to college-level math: Early outcomes using the Virginia placement test. CCRC Research Brief.
Scott-Clayton, J. (2012). Do high-stakes placement exams predict college success? CCRC working paper no. 41. Community College Research Center, Columbia University.
Thornton, D., Case, J., & Peppers, C. (2019). Low-stakes mathematics placement and preparation using EdReady. *Journal of the National College Testing Association*, 3(1), 1–9.