The University of Texas at Austin

Education Research Center

www.TexasERC.utexas.edu

POLICY BRIEF

Reforming Assessment into Developmental Education and Building the Research Base

Dan Cullinan and Dorota Rizik
MDRC
March 2023

Executive Summary

The approach to developmental education has been evolving in recent years and many colleges are implementing bold reforms, for example no longer requiring developmental education courses or changing how students are placed into courses. Some colleges are moving away from standardized tests to assess students' college readiness and instead are using alternative measures of students' performance. In doing so, it is important to understand which measure or combinations of measures are best at predicting students' success in college-level courses. This policy brief describes how data from the Texas Education Research Center (ERC) and supplemental data from seven participating Texas colleges (The University of Texas at Arlington; Southwest Texas Junior College; Texas Southern University; Texas A&M Texarkana; Lee College; El Paso Community College; Alamo Colleges District) were used to examine various measures as possible predictors of success in college-level English or math. The findings suggest that using all available measures is best, but among the simpler models, high school GPA seems to be the single best predictor of successful completion of college-level courses in English and math.

What We Studied

Colleges and universities are rethinking their approach to developmental education and implementing bold reforms.[1] Some states no longer require developmental education courses, while others are encouraging colleges to enroll students with developmental needs directly into college-level courses with corequisite supports. Colleges have also been changing how they place students into courses, with many moving away from standardized tests to assess students' college readiness and instead using alternative measures of students' performance, such as high school grades or GPA.[2] The COVID-19 pandemic has further upended traditional placement practices with the cancellation of standardized tests, leaving colleges to find other ways to assess students' college readiness (often with limited evidence about which placement practices work best for whom).[3] A long-overdue call for more racial equity further underscores the dearth of knowledge about which practices work best to promote the success of students of color and low-income students, who have long been overrepresented in developmental education.[4] State and college leaders are seeking more information and assistance in reforming their placement practices to create more equitable outcomes for all students.

The Center for the Analysis of Postsecondary Readiness (CAPR), a U.S. Department of Education Institute of Education Sciences (IES)-funded national Research and Development Center, has made strong progress in establishing rigorous evidence about which reforms are effective in improving students' college success. Led by MDRC and the Community College Research Center (CCRC), CAPR's work has revealed that important practices, such as the use of multiple measures assessment (MMA), can increase the number of students placing into and succeeding in college-level courses.[5] However, despite the strong evidence, MMA and its best practices for implementation have yet to fully penetrate the field. To advance effective practice with MMA systems and increase the knowledge base around the efficacy and equity of MMA practices, the THECB partnered with CAPR to determine which measures are most predictive for student success in college-level math and English courses in Texas.

Purpose of the Study and Research Questions

CAPR's work with ERC and supplemental data will seek to provide THECB and Texas institutions with useful information on multiple measures for MMA practices, guided by the following research questions:

1. What are the rates of developmental, co-requisite, and credit-bearing course placement across institutions in Texas?
2. Which measures are most predictive of students' successful college-level course completion in math and English?
3. Are the measures equally predictive of successful college-level course completion across different student groups?
4. Which MMA practices or models show the most promise given existing costs and implementation requirements?

How We Analyzed the Data

To answer Research Question 1 CAPR researchers used student-level course data from the ERC to observe how many (and which) students are placed into developmental, co-requisite, and credit-bearing courses across colleges, and the academic outcomes for students placed in those courses. This analysis included descriptive statistics and graphical assessments to compare colleges with different placement practices. These analyses were provided to the seven participating data colleges.

With additional student data provided by participating colleges, CAPR researchers answered Research Question 2 by using a series of predictive models built using classification methods to predict students' success in college-level credit-bearing courses. CAPR researchers focused on existing placement measures (for example, TASP/TSI) and those measures that have proven to be effective in other states (for example, high school GPA). The samples for these predictive analyses included all students at the seven data colleges who enrolled in college-level math in last five years (16,391 students), and in college-level English in last five years (58,541 students). Table 1 summarizes the demographic characteristics of the two samples.

Table 1: Characteristics of the Analysis Sample	College-Level English Enrollees	College-Level Math Enrollees
Age (%)		
20 or younger	53.2	54.0
21-30	42.5	42.9
31 or older	4.3	3.1
Gender (%)		
Men	44.3	44.7
Women	55.7	55.3
Race/Ethnicity (%)		
Asian	3.6	4.3
Black	6.9	7.4
Hispanic	68.1	60.4
Other	2.9	3.4
White	18.5	24.5

The predictive models included completion of college-level math and completion of college-level English as the outcomes of interest (respectively for the two samples), and included standardized test scores, high school GPA, and other high school metrics as predictors. Table 2 summarizes the predictors used in the models.

Table 2: Descriptive Statistics of Key Predictors	Mean	Min	Max	Percent Missing
High School Performance				
High School GPA	0.8	0	2.5	32
Number of English courses taken	5	0	17	25.2
Number of math courses taken	5.1	0	19	25.2
High School Attendance				
Days absent	7.2	0	173	25.2
Days present	150.8	1	435	25.2
Indicators of High School Course Completion (%)				
Ever passed Algebra 1	30.3	0	100	
Ever passed Algebra 2	56.7	0	100	
Ever passed English 1	34.7	0	100	
Ever passed English 2	48.6	0	100	
Ever passed English 3	46.6	0	100	
Ever passed English 4	55.1	0	100	
Ever passed Geometry	45.8	0	100	
Ever passed Pre-Calculus	39.6	0	100	

Ever passed Reading 1	2.7	0	100
Ever passed Reading 2	1.5	0	100
Ever passed Reading 3	0.6	0	100
Indicators of Completed High School Endorsements (%)			
Arts & Humanities	9.8	0	100
Business & Industry	7	0	100
Multi-Disciplinary Studies	21.9	0	100
Public Service	7.8	0	100
STEM	10.1	0	100

For Research Question 3, CAPR researchers ran the same models used to answer the second research question but using different samples of students, focused on commonly underserved groupings (for example, gender, age, and race/ethnicity). This work focused on ensuring that MMA practices are equitable and benefit all students.

For Research Question 4, CAPR researchers will analyze implementation and cost data collected directly from participating colleges (for example, quantitative information from budgets and reports of staff time spent implementing MMA systems, and qualitative information from faculty and staff interviews). A cost analysis will be performed to assess the resources required to implement and scale existing MMA practices. Results from the cost analysis will be presented in the final report.

What We Discovered

All classification models were assessed based on the Area Under the Curve (AUC) of their Receiver Operating Characteristic (ROC) curve. The ROC curve is a graph showing the performance of a classification model across all classification thresholds, and the AUC is a measure of the classifier's ability to distinguish between classes and is a summary of the ROC curve. The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes. The AUC typically ranges between 0.5 (random chance) and 1.0 (perfect prediction). Table 3 summarizes the AUC ROC values for all models.

The “kitchen sink” model with all available predictors had the highest AUC value for both the math and English samples, which was expected because this was the model with the most information. The model with all available standardized tests and high school GPA had the second highest AUC values for both samples. However, high school GPA on its own was not far behind the top performing models (with AUC values of 0.65 for English and 0.66 for math). Moreover, including information about which high school the student graduated from along with GPA improved the model minimally for both subjects, which suggests high school GPA is a strong predictor regardless of the high school context.

Table 3: AUC ROC Values by Predictor Set	Subject	
	English	Math
Kitchen Sink Models		
All available predictors	0.699	0.705
All available tests + GPA	0.683	0.690
All available tests	0.651	0.644

High School Performance		
Attendance	0.602	0.633
Completed endorsements	0.534	0.541
Completed math/English courses	0.591	0.588
High school GPA	0.650	0.659
High school student graduated from	0.582	0.587
High school student graduated from + GPA	0.659	0.661
Number of math/English courses taken	0.545	0.561
ACT		
Math		0.519
English	0.517	
Reading	0.502	
All ACT tests	0.517	0.522
All ACT tests + High School GPA	0.659	0.663
SAT		
Math		0.552
Reading	0.502	
Writing	0.535	
All SAT tests	0.553	0.562
All SAT tests + GPA	0.665	0.672
STAAR		
Algebra 1		0.535
English 1	0.558	
English 2	0.576	
All STAAR tests	0.588	0.598
All STAAR tests + GPA	0.664	0.662
TSI		
Math		0.566
Reading	0.571	
Writing	0.566	
All TSI tests	0.608	0.572
All TSI tests + GPA	0.617	0.583

Overall, the range of AUC values was small across all predictors sets, but high school GPA on its own performed better than any standardized test on its own. In fact, high school GPA alone is a better predictor of college-level success in math than all standardized tests combined. That said, the model with all available standardized tests did perform as well as GPA for the English sample. This improved performance was likely related to the amount of available data. Table 4 summarizes the percentage of each sample that had a score for more than one test compared with the percentage of each sample that had a GPA. In both the math and English samples, more students had observable predictors available in the model using all standardized tests than in the model using GPA only because more students had test scores from multiple tests than had GPA. In a sample with no missing high school GPA, standardized tests would lose this advantage.

Table 4: Percentage of Sample with Data	Subject	
	English	Math
More than one standardized test	82%	73%
High school GPA	68%	61%

Table 5 summarizes the AUC ROC values for the best performing model for each subgroup. Across all subgroups, as with the full samples, the best performing model was the "kitchen sink" model that included all available predictors. The subgroup findings suggest the models perform similarly for different groups of students.

Table 5: AUC for the Best Performing Model by Subgroup	AUC ROC by Subject	
	English	Math
Race/Ethnicity		
White	0.653	0.656
Black	0.693	0.611
Hispanic	0.684	0.693
Other	0.683	0.678
Gender		
Men	0.691	0.677
Women	0.714	0.688
Age		
Age 21+	0.709	0.647
Age 0-20	0.736	0.744

Discussion/Policy Recommendations

CAPR researchers found that the predictive utility of placement measures is similar in Texas to that in other states and systems. In general, high school GPA tends to be the best single observable predictor of success in college-level math and English courses without additional supports. Using multiple measures in addition to high school GPA marginally improves those predictions.

When considering implementing similar models to those discussed here, institutions should remember that these models provide information about how likely students are to perform well in gatekeeper courses without additional supports. Furthermore, institutions should consider the tradeoff between higher cut-offs on placement measures and accuracy of placement. For example, a cut-off of 2.5 for high school

GPA could result in higher accuracy compared with a 3.0 cut-off, but the rate of "false positive" placements would increase with a lower cut-off. That said, the simplest models using only GPA, or a combination of GPA and other standardized tests, perform almost as well as the most elaborate models, at a much lower cost.

References

1. Zachry Rutschow, E., & Mayer, A. (2018). Early findings from a national survey of developmental education practices. New York, NY: Center for the Analysis of Postsecondary Readiness, MDRC.
2. Zachry Rutschow, E., Cormier, M. S., Dukes, D., & Cruz Zamora, D. E. (2019). The changing landscape of developmental education practices: Findings from a national survey and interviews with postsecondary institutions. New York, NY: Center for the Analysis of Postsecondary Readiness, MDRC.
3. Bickerstaff, S., Kopko, E., Lewy, E. B., Raufman, J., & Zachry Rutschow, E. (2021). Implementing and scaling multiple measures assessment in the context of COVID-19.New York, NY: Center for the Analysis of Postsecondary Readiness, MDRC.
4. Ngo, F., & Kwon, W. W. (2014). Using multiple measures to make math placement decisions: Implications for access and success in community colleges. Research in Higher Education, 56(5), 442-470.
5. Barnett, E. A., Bergman, P., Kopko, E., Reddy, V., Belfield, C., & Roy, S. (2018). Multiple measures placement using data analytics: An implementation and early impacts report. New York, NY: MDRC.