INCREASING GATEKEEPER COURSE COMPLETION

Three-Semester Findings from an Experimental Study of Multiple Measures Assessment and Placement

Dan Cullinan
Dorota Biedzio

DECEMBER 2021

mdrc
BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

CCRC COMMUNITY COLLEGE RESEARCH CENTER
TEACHERS COLLEGE, COLUMBIA UNIVERSITY

Increasing Gatekeeper Course Completion

Three-Semester Findings from an Experimental Study of Multiple Measures Assessment and Placement

Dan Cullinan
Dorota Biedzio

mdrc
BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

CCRC COMMUNITY COLLEGE RESEARCH CENTER
TEACHERS COLLEGE, COLUMBIA UNIVERSITY

DECEMBER 2021

FUNDERS

Funding for this report came from the Ascendium Education Group.

Dissemination of MDRC publications is supported by the following organizations and individuals that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, Arnold Ventures, Charles and Lynn Schusterman Family Foundation, The Edna McConnell Clark Foundation, Ford Foundation, The George Gund Foundation, Daniel and Corinne Goldman, The Harry and Jeanette Weinberg Foundation, Inc., The JPB Foundation, The Joyce Foundation, The Kresge Foundation, and Sandler Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

For information about MDRC and copies of our publications, see our website: www.mdrc.org.

OVERVIEW

Colleges throughout the United States are evaluating the effectiveness of their strategies to place entering students into college-level or developmental education courses. Developmental, or remedial, courses are designed to advance the reading, writing, and math skills of students who are deemed academically underprepared for college-level courses. Placements have traditionally been determined through standardized placement testing; however, through evaluating additional types of placement tests, high school transcripts, and evaluations of student motivation, multiple measures assessments (MMAs) are becoming an increasingly popular tool to place students with greater nuance.

There is no single, correct way to design and implement a multiple measures assessment to improve course placements. Colleges must decide what measures to include, and how to combine them. This study was developed to add to the understanding of the implementation, cost, and efficacy of an MMA system using locally determined rules. As part of a randomized controlled trial, the study team evaluated MMA programs and observed 17,203 student performances across five colleges in Minnesota and Wisconsin over the course of the fall 2018, spring 2019, and fall 2019 semesters.

Findings

Across the five colleges in the random assignment study, about 15 percent of all students who were observed were placed in an alternative course level as a result of the implementation of multiple measures assessments. In this main analysis sample for whom MMA impacted their course placement, there were 1,814 students who had low test scores in English and 2,082 who had low test scores in math but who had strong high school grade point averages (GPAs) or noncognitive scores and were in the "bump-up zone."

Regarding the quantitative findings over the three-semester period:

- Program group students in the bump-up zone enrolled in more college-level courses than control group students (30.2 percentage points more in English and 19.2 percentage points more in math).
- Students in the bump-up zone who were placed into college-level English were 16 percentage points more likely to have completed the course by the end of their third college semester than their control group counterparts.
- Students in the bump-up zone who were placed into college-level math were 11 percentage points more likely to have completed the course by the end of their third college semester compared with their control group counterparts.
- Overall, all subgroups of students benefited from multiple measures placement, and MMA generally has positive impact estimates on enrollment and completion of gatekeeper courses in English and math.
- This implementation effort cost the colleges about $33 per student who went through the placement process during the three semesters of the study.

CONTENTS

CONTENTS *(continued)*

LIST OF EXHIBITS

TABLE

LIST OF EXHIBITS *(continued)*

TABLE

FIGURE

BOX

ACKNOWLEDGMENTS

The authors are thankful to the many administrators and faculty and staff members who helped implement and evaluate new ways of placing students at the five colleges participating in the randomized controlled trial: Anoka Ramsey Community College, Century College, Madison College, Minneapolis Community and Technical College, and Normandale Community College. We would also like to thank the Minnesota State and Wisconsin Technical College Systems for their cooperation and participation in this project. Thanks to Amy Kerwin and Sue Cui at Ascendium Education Group for their ideas and insight throughout the life of this project, and to Ascendium Education Group for its generous financial support of this project. We would like to thank current and recent members of the Multiple Measures Assessment team from MDRC and the Community College Research Center, including Beth Kopko, Stanley Dai, and Parker Cellura. Thanks also to our senior advisers and reviewers—Elisabeth Barnett, Rob Ivry, Michael Weiss, and Clive Belfield—for their careful reading and thoughtful feedback during the review process. We thank Rebecca Bender for editing this report and Carolyn Thomas for preparing it for publication.

The Authors

EXECUTIVE SUMMARY

Colleges throughout the United States are evaluating the effectiveness of their strategies to place entering students into college-level or developmental education courses. Developmental, or remedial, courses are designed to advance the reading, writing, and math skills of students who are deemed academically underprepared for college-level courses. This determination is usually made using standardized placement tests such as the ACCUPLACER.[1]

For years, colleges have used a single placement test, but many schools are now using multiple measures assessment (MMA)—factoring in additional test scores, high school transcripts, and evaluations of noncognitive skills—to assess and place incoming students. This practice has accelerated in the last few years, especially since the onset of the COVID-19 pandemic, when colleges looked for more flexible placement methods that were not based solely on a single, sometimes difficult to administer, test. MMA systems, like those studied in this report, are now used in states and colleges around the country.[2]

Despite the promise of MMA, millions of students each year are still being enrolled into developmental classes in math and/or English.[3] Not only does this delay students' entry into credit-bearing coursework, but those who begin their studies in developmental classes are also less likely to graduate.[4] Using MMA could be particularly significant for students of color, who are overrepresented in developmental courses.[5] MMA can improve outcomes for these students, and may help close achievement gaps.

The findings in this report are derived from a research project undertaken by MDRC and the Community College Research Center to study the use of MMA at Minnesota and Wisconsin community colleges, with funding from the Ascendium Education Group. Five colleges participated in

1. Elizabeth Zachry Rutschow, Maria S. Cormier, Dominique Dukes, and Diana E. Cruz Zamora, *The Changing Landscape of Developmental Education Practices: Findings from a National Survey and Interviews with Postsecondary Institutions (New York: Community College Research Center, Teachers College, Columbia University; and Center for the Analysis of Postsecondary Readiness, MDRC, 2019).*

2. Susan Bickerstaff, Elizabeth Kopko, Erika B. Lewy, Julia Raufman, and Elizabeth Zachry Rutschow, *Implementing and Scaling Multiple Measures Assessment in the Context of COVID-19* (New York: MDRC, 2021).

3. Xianglei Chen, Michael A. Duprey, Nichole Smith Ritchie, Lesa R. Caves, Daniel J. Pratt, David H. Wilson, Frederick S. Brown, and Katherine Leu, *High School Longitudinal Study of 2009 (HSLS:09): A First Look at the Postsecondary Transcripts and Student Financial Aid Records of Fall 2009 Ninth-Graders (Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2020).*

4. This could be for a number of reasons, including less-prepared students entering developmental courses, or because the courses themselves present an obstacle to students.

5. Xianglei Chen, Lesa R. Caves, Joshua Pretlow, Samuel Austin Caperton, Michael Bryan, and Darryl Cooney, *Courses Taken, Credits Earned, and Time to Degree: A First Look at the Postsecondary Transcripts of 2011–12 Beginning Postsecondary Students (Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2020).*

the randomized controlled trial, which compared students who were placed using the college's existing procedures (the control group) with students who were placed using MMA (the program group). The control group was placed using ACCUPLACER tests, while the program group was placed using MMA systems that incorporated high school grade point average (GPA) and noncognitive assessments, either the Learning and Study Strategies Inventory (LASSI) or the Grit test. Colleges wanted to incorporate noncognitive assessments because they believe success is not determined by content knowledge—the focus of standardized tests—alone.

This report examines the impacts of MMA on math and English gatekeeper course completion and college-level credit accumulation three semesters after students' initial placement. This report also analyzes the predictive utility of high school GPA, placement tests, and the noncognitive assessments used by the study colleges. Finally, the report provides a cost and cost-effectiveness analysis of MMA as implemented by these colleges. The primary research questions are these:

What is the effect of using multiple measures placement on the following outcomes?

- Completion of the first English college-level course (C or higher) within three semesters
- Completion of the first math college-level course (C or higher) within three semesters
- Cumulative college-level credit accumulation within three semesters

How well does each noncognitive assessment used by the participating colleges predict college course completion and persistence in the following circumstances?

- When used alone
- When used in combination with high school GPA

What was the total cost of the resources required to build and scale these MMA systems, including, where applicable, a breakdown by who incurred which costs?

What was the incremental cost per additional credit earned as a result of the MMA systems?

Measures Used and Placement Approach

All colleges in the study included the following measures in their MMA systems: placement test scores, high school GPA, noncognitive assessment results, and scores from the ACT and SAT. The specific measures and decision rules used at each college are displayed in Table ES.1.

Once the colleges selected their assessment measures, they had to decide how those measures would be combined. This was usually done by developing a set of decision rules in which each measure would be considered in a specific order to determine which classes students were eligible to take. The colleges in the study sought to automate this process as much as possible. The third column in Table ES.1 shows the sequence in which colleges considered these measures. Typically, colleges considered waivers first to identify students who would be exempt from consideration of other

TABLE ES.1 MMA Approaches at Colleges in the Multiple Measures Assessment Study

COLLEGE NAME AND STATE	TYPE OF PLACEMENT SYSTEM	MMA APPROACH AND ORDER OF STEPS	NONCOGNITIVE ASSESSMENT	COLLEGE-READY HIGH SCHOOL GPA LEVEL
Anoka-Ramsey Community College, Minnesota	Decision rule	1. Exemptions (AP/IB, ACT, SAT, MCA scores) 2. ACCUPLACER (exemption) 3. GPA or LASSI	LASSI (motivation): 50th percentile	English/Math: ≥3.0 GPA
Century College, Minnesota	Decision rule	1. Exemptions (AP/IB, ACT, SAT, MCA scores) 2. ACCUPLACER (exemption) 3. GPA or LASSI	LASSI (motivation): 50th percentile	English/Math: ≥3.0 GPA
Madison College, Wisconsin	Decision band	1. Exemption (ACT scores) 2. ACCUPLACER (decision band) 3. GPA or Grit	Grit Scale: 4+	English/Math: ≥2.6 GPA
Minneapolis Community and Technical College, Minnesota	Decision band	1. Exemptions (ACT, IB, SAT, MCA scores; college credit) 2. ACCUPLACER (decision band) 3. GPA or LASSI	LASSI (motivation): 75th percentile	English: ≥2.3 GPA Reading: ≥2.4 GPA Math: ≥3.0 GPA
Normandale Community College, Minnesota	Decision rule	1. Exemptions (AP, ACT, SAT, MCA scores; college credit) 2. LASSI 3. GPA or ACCUPLACER (exemption)	LASSI (motivation): 75th percentile	English/Reading: ≥2.5 GPA Math: ≥2.7 GPA

NOTES: Decision rules are a sequence of rules that compares each selected measure with a threshold in a predetermined order. If the threshold is met, a placement is generated; if not, another rule is applied. Decision bands are decision rules that apply only to students who fall within a certain range on a specified indicator (such as high school GPA or a placement test score), usually just below the cutoff.

GPA = grade point average, MCA = Minnesota Comprehensive Assessment, LASSI = Learning and Study Strategies Inventory.

measures. Subsequently, the results of the ACCUPLACER placement test, the high school GPA, and the noncognitive assessment would be considered. In some cases, a system of "decision bands," applicable to students within a particular score range, was used. In these cases, students who earned test scores within a certain range would be evaluated using other measures.

Identifying, Recruiting, and Randomly Assigning Students

Five colleges participated in the randomized controlled trial, including all students taking placement tests for enrollment in the fall 2018, spring 2019, and fall 2019 semesters, making three cohorts. The colleges were Anoka Ramsey Community College, Century College, Minneapolis Community and Technical College, and Normandale Community College, all in Minnesota, and Madison College in Wisconsin. Colleges chose not to include dual-enrollment students taking courses at the college while still in high school, as well as English language learners (ELLs). Dual-enrollment students come directly from high school and might go through a different placement process, and high school GPAs based on ELL coursework might have different predictive value for college coursework. Across the four Minnesota colleges, a total of 13,610 students participated in the study. The fifth college, Madison, enrolled 3,593 students, bringing the total number of randomized students to 17,203.[6] There were 12,046 students testing for English placements and 15,002 testing for math.

All 17,203 students in the sample were randomly assigned to one of two study groups. The program group placed using MMA—specifically high school GPA, noncognitive LASSI or Grit test scores, and the traditional ACCUPLACER placement test. The control group used only the single ACCUPLACER test.[7] Most of the students' placement was not changed by MMA; about 85 percent of all students were referred to the same course level regardless of the placement procedure that was used. For these students, whose placement was unchanged, the expectation is that the use of multiple measures will have no effect on their academic outcomes. For this reason, this report focuses on the main analysis sample of students whose placement was changed by MMA (or whose placement would have been changed had they been in the program group). Students in the main analysis sample were "bumped up" by MMA, so the main analysis sample is also referred to as "students in the bump-up zone." There were 1,814 students who had low test scores in English and 2,082 who had low test scores in math but who had strong high school GPAs or noncognitive scores and were bumped up.

Effects of Multiple Measures Assessment

This section presents findings on the MMA placements' estimated effects on the academic outcomes of all cohorts of students in the bump-up zone. After three semesters, it is likely that most students who were initially placed into developmental courses could have had an opportunity to take college-level courses; this allowed the research team to examine how students from the different referral groups did academically and to assess whether offering college-course placements through MMA led to higher rates of college-level course completion and credit accumulation over time. Impact estimates are summarized in Tables ES.2 and ES.3.

6. Madison randomized a large number of students, but because of implementation bottlenecks associated with a lack of automation in their placement process, only a small number of students were given the opportunity to be placed using multiple measures. This college also used different placement tests and noncognitive assessments compared with Minnesota. For these reasons, an exploratory subgroup analysis examined if there were differential effects of MMA by state.

7. The program-to-control random assignment ratio was 70:30 at Century, Minneapolis, and Madison and 50:50 at Anoka-Ramsey and Normandale, but the latter school changed the ratio to 70:30 for the fall 2019 cohort.

TABLE ES.2 Academic Outcomes After Three Semesters Among Students in the English Bump-Up Zone

OUTCOME	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL LOWER BOUND	90% CONFIDENCE INTERVAL UPPER BOUND	P-VALUE
First-semester placement						
Gatekeeper (%)	100.0	0.0	100.0	100.0	100.0	0.000
Developmental (%)	0.0	100.0	-100.0	-100.0	-100.0	0.000
Three-semester outcomes						
Gatekeeper (%)						
Enrolled	63.3	33.1	30.2	26.6	33.7	0.000
Completed (C or higher)	42.8	26.6	16.3	12.6	19.9	0.000
Failed	12.1	3.3	8.7	6.5	11.0	0.000
Withdrew	8.3	2.9	5.4	3.5	7.4	0.000
Developmental (%)						
Enrolled	7.4	42.0	-34.6	-37.4	-31.8	0.000
Completed (C or higher)	5.4	34.0	-28.6	-31.2	-25.9	0.000
Failed	1.1	5.6	-4.5	-5.8	-3.2	0.000
Withdrew	1.4	3.2	-1.8	-2.9	-0.6	0.011
College level						
Credits earned (C or higher)	2.49	2.12	0.37	0.16	0.58	0.003
Number of courses completed	0.74	0.63	0.11	0.05	0.17	0.003
All subjects						
Enrolled during first semester (%)	81.1	77.9	3.1	0.7	5.6	0.033
Enrolled during second semester (%)	66.6	67.0	-0.3	-3.9	3.3	0.887
Enrolled during third semester (%)	47.6	49.1	-1.4	-5.4	2.5	0.548
Number of semesters enrolled	1.95	1.94	0.01	-0.06	0.09	0.767
Total credits attempted	22.33	21.62	0.71	-0.32	1.75	0.258
Total credits earned	16.55	16.90	-0.34	-1.43	0.74	0.604
College-level credits earned (C or higher)	14.35	13.09	1.26	0.26	2.26	0.038
Developmental credits earned	1.06	2.91	-1.85	-2.11	-1.59	0.000
College-level courses completed	4.78	4.46	0.32	0.00	0.65	0.103
Sample size (total = 1,814)	1,126	688				

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.
Distributions may not add to 100 percent because categories are not mutually exclusive.
The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

TABLE ES.3 Academic Outcomes After Three Semesters Among Students in the Math Bump-Up Zone

OUTCOME	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL LOWER BOUND	90% CONFIDENCE INTERVAL UPPER BOUND	P-VALUE
First-semester placement						
Gatekeeper (%)	100.0	0.0	100.0	100.0	100.0	0.000
Developmental (%)	0.0	100.0	-100.0	-100.0	-100.0	0.000
Three-semester outcomes						
Gatekeeper (%)						
Enrolled	39.8	20.6	19.2	15.9	22.5	0.000
Completed (C or higher)	25.6	14.7	11.0	8.1	13.9	0.000
Failed	4.6	2.3	2.3	0.9	3.7	0.006
Withdrew	8.7	2.9	5.9	4.0	7.7	0.000
Developmental (%)						
Enrolled	4.5	33.6	-29.1	-31.6	-26.7	0.000
Completed (C or higher)	3.7	26.4	-22.8	-25.0	-20.5	0.000
Failed	0.6	5.7	-5.1	-6.3	-4.0	0.000
Withdrew	0.8	3.5	-2.8	-3.8	-1.8	0.000
College level						
Credits earned (C or higher)	2.16	1.55	0.61	0.41	0.81	0.000
Number of courses completed	0.64	0.44	0.19	0.14	0.25	0.000
All subjects						
Enrolled during first semester (%)	84.2	84.3	-0.1	-2.0	1.8	0.917
Enrolled during second semester (%)	73.8	74.0	-0.3	-3.3	2.8	0.885
Enrolled during third semester (%)	56.6	54.6	2.0	-1.6	5.6	0.363
Number of semesters enrolled	2.15	2.13	0.02	-0.05	0.08	0.693
Total credits attempted	24.85	24.75	0.09	-0.85	1.04	0.871
Total credits earned	20.37	20.35	0.02	-0.98	1.03	0.970
College-level credits earned (C or higher)	18.62	17.14	1.48	0.51	2.44	0.012
Developmental credits earned	0.65	2.25	-1.60	-1.82	-1.38	0.000
College-level courses completed	6.04	5.63	0.41	0.10	0.71	0.027
Sample size (total = 2,082)	1,189	893				

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.

Distributions may not add to 100 percent because categories are not mutually exclusive.

The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

Summary of Findings

Program group students in the bump-up zone enrolled in more college-level courses than control group students (30.2 percentage points more in English and 19.2 percentage points more in math).

Students in the bump-up zone who were placed into college-level English were 16 percentage points more likely to have completed the course by the end of their third college semester than their control group counterparts.

Students in the bump-up zone who were placed into college-level math were 11 percentage points more likely to have completed the course by the end of their third college semester compared with their control group counterparts.

Program group students in the English bump-up zone earned 1.3 more college-level credits across all subjects, and program group students in the math bump-up zone earned 1.5 more college-level credits across all subjects.

Overall, all subgroups of students benefited from multiple measures placement, and MMA generally has positive impact estimates on enrollment in and completion of gatekeeper courses in English and math.

The predictive analysis found that GPA was the best of the available predictors of success in college-level courses. The LASSI and Grit noncognitive assessments appeared to add no predictive value above and beyond that of GPA.

Implementing MMA cost the colleges $33 per student over the business-as-usual placement process. It is comparable in per-student and per-credit-earned effects to the Encouraging Additional Summer Enrollment (EASE) informational campaign.[8] The cost could likely be lowered over time either through continued use or by tweaks to the implementation.

8. Caitlin Anzelone, Michael Weiss, and Camielle Headlam, with Xavier Alemañy, *How to Encourage College Summer Enrollment: Final Lessons from the EASE Project (New York: MDRC, 2020). MDRC's Encouraging Additional Summer Enrollment (EASE) study used behavioral insights and a financial incentive with the goal of boosting enrollment rates.*

1

Introduction and Background

Colleges throughout the United States are evaluating the effectiveness of their strategies to place entering students into college-level or developmental education courses. Developmental, or remedial, courses are designed to advance the reading, writing, and math skills of students who are deemed academically underprepared for college-level courses. This determination is usually made through the use of standardized placement tests such as the ACCUPLACER.[1]

For years, colleges have used a single placement test, but that is changing. Many schools are now using multiple measures assessment (MMA)—factoring in additional test scores, high school transcripts, and evaluations of noncognitive skills—to assess and place incoming students. This practice has accelerated in the last few years, especially since the onset of the COVID-19 pandemic, when colleges looked for more flexible placement methods that were not based solely on a single, sometimes difficult to administer, test. MMA systems, like those studied in this report, are now used in states and colleges around the country.[2]

There is now evidence from two randomized controlled trial (RCT) studies with encouraging findings on the use of MMA to help some students who would have otherwise been required to take developmental classes, to take and pass college-level courses. Box 1.1 describes MMA research in the State University of New York (SUNY) system, and how Minnesota and Wisconsin colleges designed their systems differently. Other research also suggests that MMA is a promising approach because of students' increased placement into college-level courses and improved outcomes in those courses.[3]

Despite the promise of MMA, millions of students each year—about 60 percent of those entering community colleges—are still enrolling into developmental classes in math and/or English.[4] Not only does this delay students' entry into credit-bearing coursework, but those who begin their studies in developmental classes are also less likely to graduate.[5] Using MMA could be particularly significant for students of color, who are overrepresented in developmental courses.[6] MMA can improve out-

1. Rutschow, Cormier, Dukes, and Cruz Zamora (2019).
2. Bickerstaff et al. (2021).
3. Dadgar, Collins, and Schaefer (2015).
4. Chen, Duprey, et al. (2020).
5. This could be for a number of reasons, including less-prepared students entering developmental courses, or because the courses themselves present an obstacle to students.
6. Chen, Caves, et al. (2020).

BOX 1.1

SUNY Multiple Measures Assessment Study

The Center for Analysis of Postsecondary Readiness, a research collaboration between MDRC and the Community College Research Center, recently completed the first phase of a random assignment study of a multiple measures placement system that uses data analytics. The goal was to learn whether this alternative system yields placement determinations that lead to better student outcomes than a system based on test scores alone. Seven community colleges in the State University of New York (SUNY) system participated in the study. The alternative placement system used data on prior students to weight multiple measures—including placement test scores, high school grade point averages, and other measures—in predictive algorithms developed at each college. These college-specific, subject-specific algorithms were then used to place incoming students into developmental or college-level courses based on predicted probabilities of success that were compared to cutoffs or thresholds selected by college faculty and administrators. Nearly 13,000 incoming students who arrived at these colleges in the fall 2016, spring 2017, and fall 2017 terms were randomly assigned to be placed using either the status quo placement system (the business-as-usual group) or the alternative placement system (the program group). The three cohorts of students were tracked through the fall 2018 term, resulting in the collection of three to five semesters of outcomes data, depending on the cohort.

Compared to their single test–based systems, the placement algorithms used at these colleges bumped some students up from developmental to college-level courses, and bumped other students down from college-level to developmental courses. Most students' placements were unchanged. Results from the randomized controlled trial showed that students with qualifying multiple measures algorithm scores or qualifying placement test scores who were placed in developmental courses would have been more likely to pass a college-level course in math or English if they were placed directly into the college-level courses.

The multiple measures assessment (MMA) systems that were used by the Minnesota and Wisconsin colleges that are studied in the current report differ in important ways from the SUNY systems that are studied. No algorithm was run to determine weights of variables in this study, but relatively simple cutoff scores were chosen by college faculty and administrators based on ranges used in other states. No students were placed lower, or bumped down, by the MMA rules used in this study, but many were bumped up (about the same proportion as SUNY in math, but a much smaller proportion than SUNY in English). Finally, the Learning and Study Strategies Inventory and Grit noncognitive assessments were used to inform placement in this study. These were not used for placement by the SUNY study colleges.

comes for these students, and may help close achievement gaps with students who are traditionally placed into college-level courses.

There is more than one way to design and implement MMA to improve course placements. Colleges must decide which measures to include, factoring in the difficulty of obtaining certain kinds of information about students. Most often, the high school grade point average (GPA) is considered along with placement test scores because it is consistently the most predictive measure available for success in college-level courses.[7] Standardized test results such as SAT and ACT test scores and other measures such as results from noncognitive assessments may also be included.[8]

Administrators must then determine the relative importance of this information and how it is evaluated in assessing academic potential. Options range from a simple waiver system in which one or more criteria are used to waive student placement tests, to more complex methods, including using predictive models to place students based on their likelihood of success in the first-year, gatekeeper courses in English and math.

About the Study

The findings in this report are derived from a research project undertaken by MDRC and the Community College Research Center to study the use of MMA at Minnesota and Wisconsin community colleges, with funding from the Ascendium Education Group. Included in this report are results from MMA placement systems using decision rules that were developed based on prior research and local knowledge; they all incorporated noncognitive assessments. MDRC also created a guidebook describing lessons learned during implementation to help other colleges develop similar systems.[9]

Five colleges participated in the randomized controlled trial, which compared students who were placed using the college's existing procedures (the control group) with students who were placed using MMA (the program group): Anoka Ramsey Community College, Century College, Minneapolis Community and Technical College, and Normandale Community College, all in Minnesota, and Madison College in Wisconsin. The research team provided technical assistance to college staff to create MMA systems incorporating locally determined decision rules. The specific measures and decision rules used at each college are shown in Table 1.1. All five colleges took considerable effort to build systems that automated the placement process as much as possible, with an eye toward scaling it up in the future to apply to their full student populations.

Colleges in this project began enrolling students into the study in the fall of 2018 and continued to do so through the fall of 2019, for a total of three semesters. Except for students who opted out (a rare occurrence), qualifying students enrolling at each college were randomly assigned to be placed using the MMA system or their college's traditional, "business as usual" placement system, typically using the ACCUPLACER placement test alone. Student outcomes in the two groups were compared three semesters following their placement. This follow-up period allowed time for students who were placed into developmental courses to finish them, enroll in the gatekeeper course, and finish it. This sequence would take at least two semesters. Colleges used simpler MMA systems that were

7. Belfield and Crosta (2012); Scott-Clayton (2012); Barnett et al. (2018).

8. Noncognitive assessments measure student qualities, characteristics, and attitudes, apart from content knowledge that may influence success in educational endeavors. Since these assessments require cognition, some people prefer other terms such as "nonacademic," "soft skill," or "21st-century skills assessments."

9. Cullinan et al. (2018).

TABLE 1.1 MMA Approaches at Colleges in the Multiple Measures Assessment Study

COLLEGE NAME AND STATE	TYPE OF PLACEMENT SYSTEM	MMA APPROACH AND ORDER OF STEPS	NONCOGNITIVE ASSESSMENT	COLLEGE-READY HIGH SCHOOL GPA LEVEL
Anoka-Ramsey Community College, Minnesota	Decision rule	1. Exemptions (AP/IB, ACT, SAT, MCA scores) 2. ACCUPLACER (exemption) 3. GPA or LASSI	LASSI (motivation): 50th percentile	English/Math: ≥3.0 GPA
Century College, Minnesota	Decision rule	1. Exemptions (AP/IB, ACT, SAT, MCA scores) 2. ACCUPLACER (exemption) 3. GPA or LASSI	LASSI (motivation): 50th percentile	English/Math: ≥3.0 GPA
Madison College, Wisconsin	Decision band	1. Exemption (ACT scores) 2. ACCUPLACER (decision band) 3. GPA or Grit	Grit Scale: 4+	English/Math: ≥2.6 GPA
Minneapolis Community and Technical College, Minnesota	Decision band	1. Exemptions (ACT, IB, SAT, MCA scores; college credit) 2. ACCUPLACER (decision band) 3. GPA or LASSI	LASSI (motivation): 75th percentile	English: ≥2.3 GPA Reading: ≥2.4 GPA Math: ≥3.0 GPA
Normandale Community College, Minnesota	Decision rule	1. Exemptions (AP, ACT, SAT, MCA scores; college credit) 2. LASSI 3. GPA or ACCUPLACER (exemption)	LASSI (motivation): 75th percentile	English/Reading: ≥2.5 GPA Math: ≥2.7 GPA

NOTES: Decision rules are a sequence of rules that compares each selected measure with a threshold in a predetermined order. If the threshold is met, a placement is generated; if not, another rule is applied. Decision bands are decision rules that apply only to students who fall within a certain range on a specified indicator (such as high school GPA or a placement test score), usually just below the cutoff.
GPA = grade point average, MCA = Minnesota Comprehensive Assessment, LASSI = Learning and Study Strategies Inventory.

not directly dependent on predictive models, such as those in the SUNY study.[10] In addition, these colleges used noncognitive assessments, either the Learning and Study Strategies Inventory (LASSI) or the Grit test, with the understanding that college success is not determined by content knowledge —the focus of standardized tests—alone (see Box 1.2).

BOX 1.2

Noncognitive Assessments

Noncognitive assessments can be valuable sources of information about students' readiness for college and may be particularly useful in cases where high school transcript data are unavailable or for students who have been out of the education system for an extended time. However, very little information is available about whether existing noncognitive assessments are useful in making placement decisions. The study also provides information on their value in creating effective multiple measures assessment systems.

The Grit Scale was selected by one of the study colleges, while the Learning and Study Strategies Inventory (LASSI) was used by four colleges. Before the evaluation was launched, colleges reviewed research on several noncognitive assessments to understand the extent to which each one predicted the successful completion of college-level courses as well as the time students would spend in testing and the cost of the assessment options.* The Grit Scale measures perseverance and passion for long-term goals. It is available at no cost and has been shown to predict positive outcomes in college settings.† The LASSI is a much longer assessment that addresses factors ranging from motivation to comfort with testing. Some colleges appreciated the opportunity to have more extensive information about their incoming students, despite the cost to use the test and the greater amount of time students spent in testing. For placement purposes, colleges used only the LASSI's motivation scale, which prior research shows is predictive of success in college.‡

NOTES: *See Cullinan et al. (2018) for more information on different noncognitive test options.
†Duckworth, Peterson, Matthews, and Kelly (2007).
‡Carson (2012); Rugsaken, Robertson, and Jones (1998).

The current study was designed to improve the knowledge base on the implementation, cost, and efficacy of an MMA system that uses locally determined rules. An earlier report by MDRC addressed questions about the implementation of MMA at the five RCT colleges and one additional college that was not in the trial study.[11] It also presented the short-term impacts of using MMA to "bump up" students into college-level gatekeeper classes based on MMA results, including enrollment and pass rates, in the first semester after placement testing at four of those colleges. The "bump-up zone" is the range of high school GPA or noncognitive assessment scores that would allow a student with

10. Barnett, Kopko, Cullinan, and Belfield (2020).

11. Cullinan et al. (2019).

ACCUPLACER scores below the minimum traditionally required for gatekeeper course enrollment to enroll in those courses anyway.

This report examines the impacts on math and English gatekeeper course completion and college-level credit accumulation three semesters after students took the placement test, allowing time for students who were placed into developmental courses to finish them, and subsequently enroll in and finish the gatekeeper course. These are the primary outcomes of interest because they are most affected by placement systems selecting which students get to enroll in gatekeeper courses (required in many cases to take other college-level courses), and which students must take developmental courses, putatively in order to increase their probability of success in college-level courses.

This report also adds the fifth college, for which data were unavailable at the time of the first report, to the sample. The report also analyzes the predictive utility of high school GPA, placement tests, and the noncognitive assessments used by the study colleges. Finally, the report provides a cost and cost-effectiveness analysis of MMA as implemented by these colleges.

The primary research questions are these:

What is the effect of using multiple measures to bump up student placements on the following outcomes?

- Completion of the first college-level course (C or higher) within three semesters:
  - in English
  - in math

- Cumulative college-level credit accumulation within three semesters

- How well does each noncognitive assessment used by the participating colleges predict college course completion and persistence in the following circumstances?
  - When used alone
  - When used in combination with high school GPA

What is the total cost of the resources required to build and scale MMA systems, including, where applicable, a breakdown by who incurred which costs?

What is the incremental cost per additional credit earned as a result of the MMA systems?

About This Report

This report describes the development of MMA systems at the participating colleges and presents impact findings from three semesters of follow-up. Chapter 1 introduces the project. Chapter 2 describes the sample of randomized students. Chapter 3 discusses the impacts of using MMA placement on academic outcomes after three semesters. Chapter 4 examines the utility of noncognitive assessments in predicting success in college-level courses. Chapter 5 provides estimates of the cost and cost-effectiveness of these MMA systems. Chapter 6 considers the implications of this study for practice and future research.

2

Sample Intake, Sample Characteristics, and Data Sources

Five colleges participated in the randomized controlled trial, including all students taking placement tests for enrollment in the fall 2018, spring 2019, and fall 2019 semesters, making three cohorts. Colleges chose not to include dual-enrollment students taking courses at the college while still in high school, as well as English language learners (ELLs). Dual-enrollment students come directly from high school and might go through a different placement process, and high school grade point averages (GPAs) based on ELL coursework might have different predictive value for college coursework. However, one college—Normandale—did include ELL students. Across the four Minnesota colleges previously discussed in the early findings report, a total of 13,610 students participated in the study.[1] The fifth college, Madison, enrolled 3,593 students, bringing the total number of randomized students to 17,203.[2] There were 12,046 students testing for English placements and 15,002 testing for math. Students may not have had to test in both subjects if they had high enough ACT scores, Minnesota Comprehensive Assessment scores for specific subjects, or eligible transfer credits in either English or math.

Whose Placement Changed Under MMA?

All 17,203 students in the sample were randomly assigned to one of two study groups. The program group placed using MMA—specifically high school GPA, noncognitive Learning and Study Strategies Inventory (LASSI) or Grit test scores, and the traditional ACCUPLACER placement test. The control group placed using only the ACCUPLACER test, though these students also had scores on the other multiple measures, making it possible to see which of them would have been eligible for a bump up had they been in the program group.[3] Table 2.1 shows the breakdown of students from both study groups who placed into developmental courses, placed into college-level courses, or fell into a zone

1. Cullinan et al. (2019).

2. Madison randomized a large number of students, but because of implementation bottlenecks associated with a lack of automation in their placement process, only a small number of students were given the opportunity to be placed using multiple measures. This college also used different placement tests and noncognitive assessments compared with Minnesota. For these reasons, an exploratory subgroup analysis examined if there were differential effects of MMA by state.

3. Carson (2012); Duckworth, Peterson, Matthews, and Kelly (2007). The program-to-control random assignment ratio was 70:30 at Century, Minneapolis, and Madison and 50:50 at Anoka-Ramsey and Normandale, but the latter school changed the ratio to 70:30 for the fall 2019 cohort.

TABLE 2.1 Multiple Measures Placement by Subject

SUBJECT (%)	PROGRAM	CONTROL	ALL	N
English				12,046
Always developmental	37.6	34.7	36.5	4,391
Bump-up zone	15.2	14.8	15.1	1,814
Always college level	47.2	50.5	48.5	5,841
Math				15,002
Always developmental	73.5	71.3	72.6	10,894
Bump-up zone	13.1	15.1	13.9	2,082
Always college level	13.4	13.6	13.5	2,026

SOURCE: Placement data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

that resulted in higher-level course placements (the "bump-up zone"). This table shows that using MMA, 15 percent of all program students in English and 14 percent of all program students in math were eligible for placement into a college-level course rather than a developmental class. Table 2.1 also shows that a similar percentage of students in the control group would have been eligible for placement into college-level classes by the MMA rules had they been in the program group. Given that only 14 to 15 percent of students' placements changed because of MMA, most students were referred to the same course regardless of the placement procedure that was used.

The "always developmental" and "always college-level" groups represent those for whom the referral approach (MMA versus business as usual) had no effect on placement because they were referred to the same course level regardless of the referral approach. For those students whose placement was unchanged, the expectation is that the use of multiple measures will have no positive (or negative) effect on their academic outcomes. The referral approach did have an effect on placement for students in the bump-up zone—these students were eligible for college-level courses because of MMA. For this reason, the discussion of the impacts of MMA on students' academic success focuses on students in the bump-up zone, who did have a higher placement because of multiple measures placement.

There were 1,814 students who had low test scores in English and 2,082 who had low test scores in math but who had strong high school GPAs or noncognitive scores. This subset of students makes up the main analysis sample and falls into what the research team calls the "the bump-up zone"—those who would have been referred to developmental courses under the colleges' business-as-usual placement system or college-level classes under an MMA system. Within this main analysis sample, all program group students were given the opportunity to take college-level courses, while students in the control group were required to take a developmental education class first.

Characteristics of the Main Analysis Sample

The students in the main analysis sample were mostly young, female, and white. However, students of color composed a sizable portion of the sample (41.7 percent were students of color, compared with 47.5 percent who were white). This representation is important because students of color are overrepresented in developmental courses compared with college-level courses when using business-as-usual placement systems. By having a representative sample, this study can gauge if MMA can improve academic outcomes for students of color by placing more of them into college-level courses.

Table 2.2 presents demographic characteristics of students in the main analysis sample. Overall, the students in the program and control groups were similar in age, gender, and race/ethnicity after random assignment.[4] However, there were a few differences between the two groups. For example, there was a 4.2 percentage point difference between full-time enrollees in the program group and the control group (52.0 percent compared with 47.8 percent, respectively). Unlike the other variables presented in this table, this variable (and the Pell eligibility variable listed) were collected post-randomization and were likely affected by the intervention itself. Appendix Table A.1 shows the same characteristics, but for the full study sample of all randomized students, not just for the main analysis sample. Among the full sample, all characteristics are balanced between the program and control groups.

Summary of the Measures

Table 2.3 shows the main analysis sample's averages on the measures used for placement. ACCUPLACER (Classic) score averages are shown for each test among those who attempted each test, on a scale of 20 to 120.[5] Reading comprehension, sentence skills, and elementary algebra tests with scores of at least 75 are required for college-level placement under business-as-usual rules. The exact cutoffs varied by college in the study, but about half of new students typically placed into developmental courses in English and reading, while 85 to 90 percent of students placed into developmental courses in math across the five colleges.

The available high school GPA for students in the program group and the control group averaged around 3.2, with over 50 percent of students having GPAs of 3.0 or better, and only 18 to 19 percent missing a GPA. LASSI motivation scores were above 50 (out of 100) for most students who took this test, which put these students above the MMA cutoffs at the Minnesota colleges.[6] About a quarter of the students in both groups did not take the LASSI test in the colleges that administered it. Grit scores for students in the program and control groups were around 3.6 on a scale of 0 to 5, which

4. An omnibus F-test of all baseline characteristics and multiple measures found no significant differences between research groups. The random assignment procedure ensured that students who were assigned to the program group were similar to those who were assigned to the control group. Because of this, any differences in student outcomes observed between groups can be attributed to the specific placement procedure that was used.

5. The Classic is the previous version of the ACCUPLACER test. The current, widely used version is called Next Generation.

6. Weinstein, Palmer, and Acee (2016). Students who score above the 75th percentile often do not need to work on the strategies or skills for a given scale. Students who score between the 75th and the 50th percentile on any scale should consider improving the relevant learning and study skills to optimize their academic performance.

TABLE 2.2 Baseline Characteristics of Bump-Up Zone Students

CHARACTERISTIC (%)	PROGRAM GROUP	CONTROL GROUP	BOTH GROUPS
Age			
20 and under	66.9	69.1	67.8
21-30	17.0	14.9	16.2
31 and over	5.7	4.1	5.1
Age missing	10.4	11.9	11.0
Gender			
Male	33.5	33.7	33.6
Female	56.0	54.5	55.4
Gender missing	10.5	11.8	11.0
Race/ethnicity			
Asian	8.6	8.1	8.4
Black	16.0	14.2	15.3
Hispanic	10.9	9.2	10.3
White	47.5	50.2	48.6
Other	6.2	5.7	6.0
Race/ethnicity missing	10.8	12.5	11.5
Enrollment status			
Full time	52.0	47.8	50.4
Part time	30.9	31.5	31.1
Enrollment status missing	17.1	20.7	18.5
Pell eligibility			
Yes	35.6	32.8	34.5
No	45.7	47.2	46.3
Pell eligibility missing	18.7	20.0	19.2
Sample size	2,006	1,405	3,411

SOURCE: Demographic data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Distributions may not add to 100 percent because of rounding.
Enrollment status represents enrollment in the first semester. For one of the sites, this was determined based on credits attempted in the transcript data.

TABLE 2.3 Multiple Measures Assessment Scores of Bump-Up Zone Students

TEST	PROGRAM GROUP	SD	CONTROL GROUP	SD	DIFFERENCE	P-VALUE
ACCUPLACER scores[a]						
Arithmetic	47.2	25.7	47.8	23.7	-0.7	0.658
Elementary algebra	66.9	26.8	67.8	23.8	-0.9	0.362
College-level math	33.8	11.6	33.0	9.4	0.8	0.197
Reading comprehension	74.5	19.5	74.7	18.3	-0.1	0.878
Sentence skills[b]	74.2	17.4	75.8	17.8	-1.6	0.173
High school GPA (%)						0.448
3.5-4.0	20.8		20.8			
3.0-3.4	32.8		35.3			
2.5-2.9	24.2		23.9			
2.0-2.4	2.3		1.9			
1.9 or lower	0.7		0.4			
GPA missing	19.2		17.7			
LASSI score (%)						0.171
50-100	56.1		58.8			
0-49	18.4		16.1			
LASSI score missing	25.5		25.1			
Grit score	3.7	0.4	3.5	0.7	0.2	0.183
Sample size (total = 3,411)	2,006		1,405			

SOURCE: Test scores, high school GPA, and LASSI and Grit scores provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.

Statistical significance levels are indicated as: *** = 1 percent, ** = 5 percent, * = 10 percent.

The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

To assess differences between the research groups, chi-square tests were used for categorical variables and two-tailed t-tests were used for continuous variables.

SD = standard deviation, GPA = grade point average, LASSI = Learning and Study Strategies Inventory.

[a]ACCUPLACER test scores can range from 0 to 120.

[b]Only Normandale Community College used the sentence skills test to determine course placement for English. The other three Minnesota colleges used the reading comprehension test, and Madison used a combination of the two tests to determine course placement for English.

was below the MMA cutoff for this measure; most students did not have a Grit score because of the way the Grit was administered (see Box 3.1 in Chapter 3). It is important to note the missingness in GPA and the noncognitive assessments, because missingness on both these measures decreases the number of students in the bump-up zone. To be in the bump-up zone, students need to have at least one of the two measures.

There was no evidence of systematic differences between program and control groups on the placement tests, noncognitive assessments, or high school GPA at the time of placement (that is, "baseline" characteristics) in the main analysis sample. Appendix Table A.2 shows the same measures, but for the full sample of all randomized students. In the full sample of students, there were differences between the two research groups for reading comprehension, sentence skills, and grit, though the differences were small in magnitude.[7]

Data Sources and Follow-Up Periods

All analyses are based on data provided by the five colleges. These data included placement test data (including multiple measures data), college transcript records, and demographic information. Placement data were from the winter, spring, and summer of 2018 for the first cohort; the summer and fall of 2018 and winter of 2019 for the second cohort; and the winter, spring, and summer of 2019 for the third cohort. Transcript data, which contained information about courses taken and were used to calculate all key outcomes, such as enrollment, progress in math and English, credits attempted, and credits earned, were from the fall 2018 semester through the fall 2020 semester, resulting in three semesters of follow-up for all cohorts.

7. Differences in the multiple measures of the full sample had small effect sizes (less than 0.05 σ for Reading Comprehension, 0.09 σ for Sentence Skills, and 0.16 σ for Grit). There were no differences between the two research groups on the multiple measures among the main analysis sample.

3

Effects of Multiple Measures Assessment

This chapter presents findings on multiple measures assessment (MMA) placements' estimated effects on the academic outcomes of all cohorts of students in the bump-up zone. The chapter summarizes the main academic effects after students were randomly assigned to either the program group or the control group and placed into courses, and how MMA placement affected course completion and credit accumulation after three semesters (the primary outcomes for this project).[1] After three semesters, it is likely that most students who were initially placed into developmental courses could have had an opportunity to take college-level courses; this allowed the research team to examine how students from the different referral groups did academically and to assess whether offering college-course placements through MMA led to higher rates of college-level course completion and credit accumulation over time.

Summary of Findings

Program group students in the bump-up zone enrolled in more college-level gatekeeper courses than control group students (30.2 percentage points more in English and 19.2 percentage points more in math). Students in the bump-up zone who were placed into gatekeeper English were 16 percentage points more likely to have completed the course by the end of their third semester than their control group counterparts. Students in the bump-up zone who were placed into gatekeeper math were 11 percentage points more likely to have completed the course by the end of their third semester compared with their control group counterparts. Program group students in the English bump-up zone earned 1.3 more college-level credits across all subjects, and program group students in the math bump-up zone earned 1.5 more college-level credits across all subjects. Overall, all subgroups of students benefited from multiple measures placement and generally showed improvements in enrollment in and completion of gatekeeper courses in English and math.

Effects on Educational Outcomes During the First Three Semesters

Tables 3.1 and 3.2 show the academic outcomes for the main analysis sample of students who were bumped up in English and math, respectively. All students in both research groups in these tables had ACCUPLACER scores that were below the necessary cutoffs for the college-level course, which would have placed them into developmental courses under the business-as-usual placement rules.

1. Cullinan and Barnett (2021).

TABLE 3.1 Academic Outcomes After Three Semesters Among Students in the English Bump-Up Zone

OUTCOME	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL LOWER BOUND	90% CONFIDENCE INTERVAL UPPER BOUND	P-VALUE
First-semester placement						
Gatekeeper (%)	100.0	0.0	100.0	100.0	100.0	0.000
Developmental (%)	0.0	100.0	-100.0	-100.0	-100.0	0.000
Three-semester outcomes						
Gatekeeper (%)						
Enrolled	63.3	33.1	30.2	26.6	33.7	0.000
Completed (C or higher)	42.8	26.6	16.3	12.6	19.9	0.000
Failed	12.1	3.3	8.7	6.5	11.0	0.000
Withdrew	8.3	2.9	5.4	3.5	7.4	0.000
Developmental (%)						
Enrolled	7.4	42.0	-34.6	-37.4	-31.8	0.000
Completed (C or higher)	5.4	34.0	-28.6	-31.2	-25.9	0.000
Failed	1.1	5.6	-4.5	-5.8	-3.2	0.000
Withdrew	1.4	3.2	-1.8	-2.9	-0.6	0.011
College level						
Credits earned (C or higher)	2.49	2.12	0.37	0.16	0.58	0.003
Number of courses completed	0.74	0.63	0.11	0.05	0.17	0.003
All subjects						
Enrolled during first semester (%)	81.1	77.9	3.1	0.7	5.6	0.033
Enrolled during second semester (%)	66.6	67.0	-0.3	-3.9	3.3	0.887
Enrolled during third semester (%)	47.6	49.1	-1.4	-5.4	2.5	0.548
Number of semesters enrolled	1.95	1.94	0.01	-0.06	0.09	0.767
Total credits attempted	22.33	21.62	0.71	-0.32	1.75	0.258
Total credits earned	16.55	16.90	-0.34	-1.43	0.74	0.604
College-level credits earned (C or higher)	14.35	13.09	1.26	0.26	2.26	0.038
Developmental credits earned	1.06	2.91	-1.85	-2.11	-1.59	0.000
College-level courses completed	4.78	4.46	0.32	0.00	0.65	0.103
Sample size (total = 1,814)	1,126	688				

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.
Distributions may not add to 100 percent because categories are not mutually exclusive.
The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

TABLE 3.2 Academic Outcomes After Three Semesters Among Students in the Math Bump-Up Zone

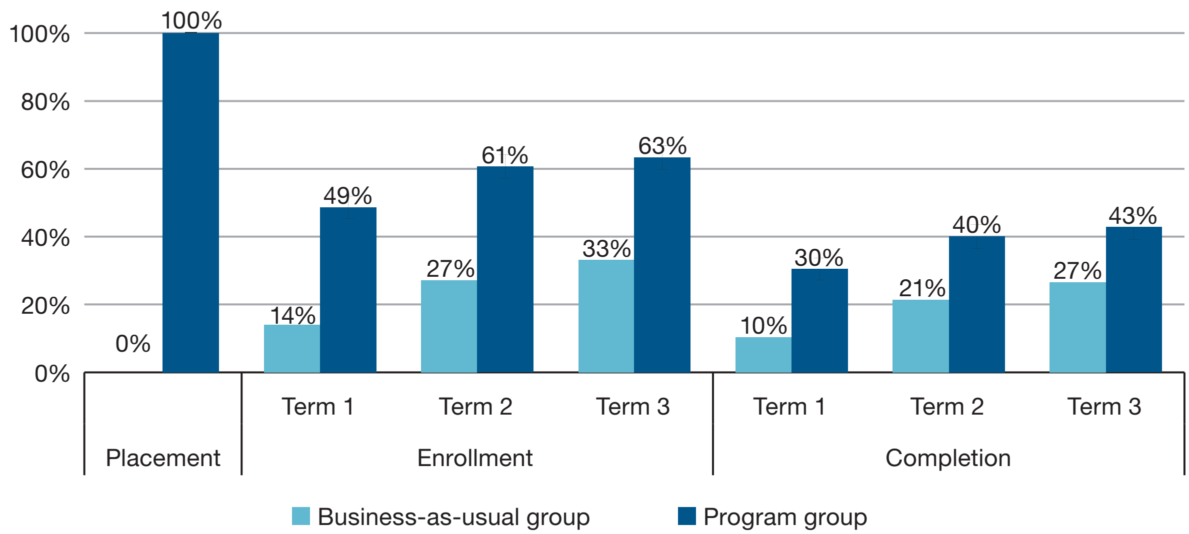
OUTCOME	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL		P-VALUE
				LOWER BOUND	UPPER BOUND	
First-semester placement						
Gatekeeper (%)	100.0	0.0	100.0	100.0	100.0	0.000
Developmental (%)	0.0	100.0	-100.0	-100.0	-100.0	0.000
Three-semester outcomes						
Gatekeeper (%)						
Enrolled	39.8	20.6	19.2	15.9	22.5	0.000
Completed (C or higher)	25.6	14.7	11.0	8.1	13.9	0.000
Failed	4.6	2.3	2.3	0.9	3.7	0.006
Withdrew	8.7	2.9	5.9	4.0	7.7	0.000
Developmental (%)						
Enrolled	4.5	33.6	-29.1	-31.6	-26.7	0.000
Completed (C or higher)	3.7	26.4	-22.8	-25.0	-20.5	0.000
Failed	0.6	5.7	-5.1	-6.3	-4.0	0.000
Withdrew	0.8	3.5	-2.8	-3.8	-1.8	0.000
College level						
Credits earned (C or higher)	2.16	1.55	0.61	0.41	0.81	0.000
Number of courses completed	0.64	0.44	0.19	0.14	0.25	0.000
All subjects						
Enrolled during first semester (%)	84.2	84.3	-0.1	-2.0	1.8	0.917
Enrolled during second semester (%)	73.8	74.0	-0.3	-3.3	2.8	0.885
Enrolled during third semester (%)	56.6	54.6	2.0	-1.6	5.6	0.363
Number of semesters enrolled	2.15	2.13	0.02	-0.05	0.08	0.693
Total credits attempted	24.85	24.75	0.09	-0.85	1.04	0.871
Total credits earned	20.37	20.35	0.02	-0.98	1.03	0.970
College-level credits earned (C or higher)	18.62	17.14	1.48	0.51	2.44	0.012
Developmental credits earned	0.65	2.25	-1.60	-1.82	-1.38	0.000
College-level courses completed	6.04	5.63	0.41	0.10	0.71	0.027
Sample size (total = 2,082)	1,189	893				

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.
Distributions may not add to 100 percent because categories are not mutually exclusive.
The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

However, all students in both research groups in these tables also had high school grade point averages (GPAs) or noncognitive scores that exceeded the MMA cutoffs at their colleges. This means that in this analysis sample, all program students were placed into college-level courses and all control students were placed into developmental courses. This is reflected in Figures 3.1 and 3.2, which show placement in, enrollment in, and completion of gatekeeper courses in English and math over time. Every program group student in the bump-up zone was placed into gatekeeper courses and every control group student in the bump-up zone was placed into developmental courses.

FIGURE 3.1 College-Level English Course Outcomes (Among Students in the English Bump-Up Zone)



SOURCE: Transcript and placement data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

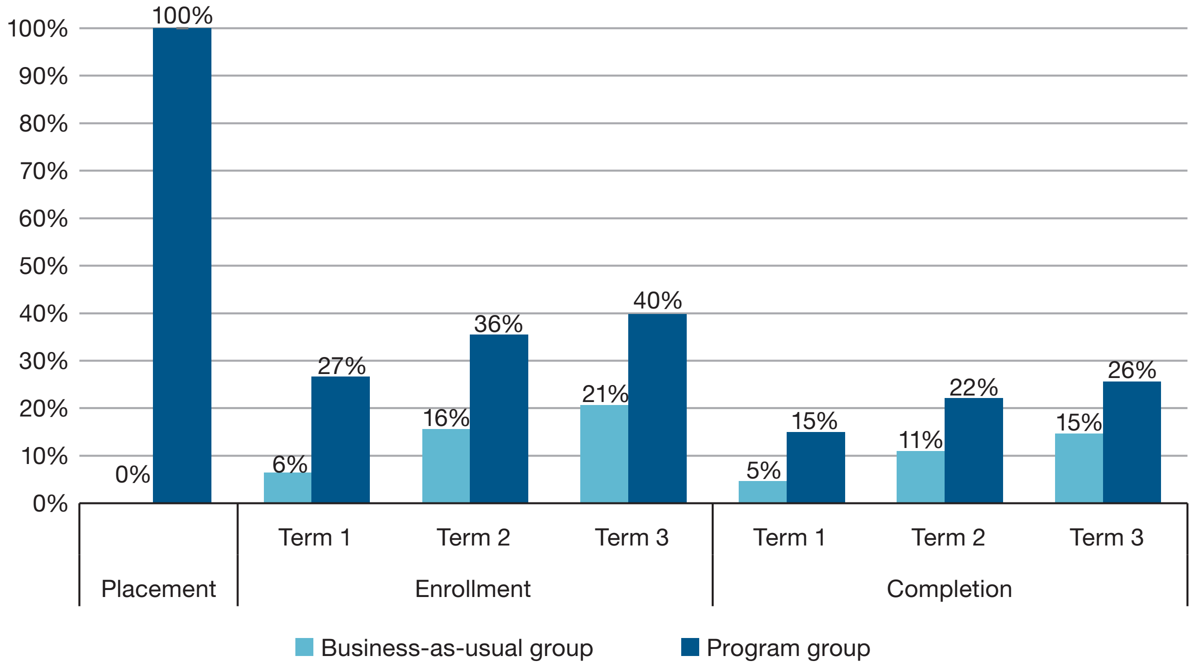
NOTES: Rounding may cause slight discrepancies in sums and differences. Distributions may not add to 100 percent

What Happened to Students Bumped Up in English?

Overall Enrollment

Students who were referred to college-level English classes instead of developmental English were more likely to enroll in college during the first semester (among students in the English bump-up zone). About 81 percent of students in the program group enrolled in any course across all subjects (developmental or college level) during the first semester, compared with 78 percent of students in the control group, a difference of 3 percentage points. This indicates that the placement into the college-level course not only affected enrollment in that subject, but these students were more likely to enroll in college in the first semester after initial placement if they were referred to college-level English. It is possible that a developmental placement itself was a barrier to students' overall enrollment because it could prevent students from enrolling in classes that most interested them. It may

FIGURE 3.2 College-Level Math Course Outcomes (Among Students in the Math Bump-Up Zone)



SOURCE: Transcript and placement data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences. Distributions may not add to 100 percent because categories are not mutually exclusive.

have been because students felt discouraged. In subsequent semesters, this difference diminished as enrollment became similar for the program and control groups.

Academic Progress Through English

Students who were referred to college-level English, instead of developmental English, were 16 percentage points more likely to complete college-level English within three semesters (among students in the English bump-up zone). This difference was likely driven by more program group students taking the college-level course compared with students in the control group. In the program group, 63 percent of students enrolled in gatekeeper English within three semesters and few of them took the developmental course (because they were not placed in it).[2] On the other hand, in the control group, only a third of students enrolled in gatekeeper English over the course of three semesters. Fewer control group students enrolled in the gatekeeper course, likely because they had to first pass the developmental course (42 percent took the developmental course within three semesters of initial

2. It is possible that some students did not follow their recommended placement—for example, if a student was placed into a gatekeeper course, but felt they were not ready to take college-level courses, they may have chosen to take the developmental course instead.

placement and 34 percent completed it). Overall, since more program students took the college-level course, more completed it.

Students who were referred to college-level English instead of developmental English also earned more college-level English credits after three semesters. However, this difference is small. Program group students earned only 0.37 more college-level English credits than the control group. It is possible that control group students may “catch up” in earning English college-level credits over time, because a third of students in the control group did eventually enroll in gatekeeper English (and 27 percent passed it) over the three semesters following their initial placement into developmental English.

Overall Academic Progress

Students who were referred to college-level English instead of developmental English earned more college-level credits across all subjects. The program group earned 1.26 more college-level credits than the control group. Conversely, the control group earned 1.85 more developmental credits than the program group. It is possible that being referred to college-level English instead of developmental English eliminated barriers that are associated with developmental education (such as feeling discouraged). It is also possible that the initial enrollment boost helped program students earn more college-level credits because they had more time to accumulate credits, earning nearly as many college-level credits as their counterparts earned developmental credits in this follow-up period.

What Happened to Students Bumped Up in Math?

Overall Enrollment

During the first three semesters, there were no differences in college enrollment between the program and control groups (among students in the math bump-up zone). Students enrolled in a similar number of semesters regardless of referral approach.

Academic Progress Through Math

Students who were referred to college-level math instead of developmental math were 11 percentage points more likely to complete college-level math within three semesters (among students in the bump-up zone). This difference was likely driven by more program group students taking the college-level course compared with students in the control group. In the program group, 40 percent of students enrolled in gatekeeper math within three semesters, and few of them took the developmental course. Among the control group, only 21 percent of students enrolled in gatekeeper math over the course of three semesters—probably because most students in the control group took the developmental course first (34 percent took the developmental course within three semesters of initial placement and 26 percent completed it). Overall, since more program students took the college-level course, more of them completed it. Students who were referred to college-level math instead of developmental math also earned 0.61 more college-level math credits after three semesters compared with the control group.

Overall Academic Progress

Students who were referred to college-level math instead of developmental math completed more college-level courses and earned more college-level credits across all subjects. The program group earned 1.48 more college-level credits than the control group. Conversely, the control group earned 1.60 more developmental credits than the program group. Given that most college-level courses are worth 3 credits, a difference of 1.48 college-level credits between the research groups has practical significance because it could suggest, for example, that half of the students in the program group passed an extra college-level course.

Differences Between English and Math

Being bumped up in either subject resulted in more enrollment in and completion of the gatekeeper course in that subject, albeit with lower base rates for enrollment in math. The differences in the completion impact estimates between the two subjects are driven by the higher enrollment in English compared with math. There was a larger impact on enrollment into college-level English in the English bump-up zone than there was on enrollment into college-level math in the math bump-up zone. This translated into larger impacts on college-level English course completion than on college-level math. It is possible that the differences in enrollment between the two subjects may have been driven by anxiety surrounding math, which is generally thought of as the more difficult subject, and more anxiety may lead to less enrollment.

Being bumped up in English increased overall college enrollment in the first semester, but there was not a significant difference in overall college enrollment caused by being bumped up to college-level math. So, overall college enrollment was more similar between the program and control groups for math. This suggests that students may be less discouraged by developmental placement in math than in English. Perhaps being an underprepared math student is perceived less negatively by these students than is being an underprepared English student, or perhaps students generally perceive math as harder. Also, course catalogs from the participating colleges indicate that gatekeeper English is required as a prerequisite for more courses than is gatekeeper math, so students placed in gatekeeper English may take more courses in other subjects, thus increasing their overall college enrollment.

Figures 3.1 and 3.2 present the differences between the program and control groups in enrollment in and completion of the gatekeeper course in English and math, respectively. These differences are presented by semester (as opposed to cumulatively after three semesters) to elucidate possible trends over time. The impacts in English decrease slightly from semester to semester, so there might be further "fade-out" as more semesters pass. On the other hand, in math, there is no fade-out in either outcome over time. So, while English enrollment and completion rates are higher than those for math, the difference between the two subjects might diminish if longer follow-up is included, perhaps because more program students take the math gatekeeper as time goes on.

Dividing the percentage of students passing a course by the percentage of students enrolling in the same course yields its pass rate. Among those in the program group who were bumped up in English, 63 percent took the college-level English course and about 43 percent passed it. This yields a 68 percent pass rate in English (43 percent out of 63 percent). The same calculation yields a math pass rate of 65 percent (26 percent out of 40 percent).[3] These pass rates may be relevant to instructors, some of whom expressed concern that MMA allowed students with lower placement test scores into their classrooms.

A representation of what might be perceived as the "status quo" pass rate can be calculated from Appendix Tables A.3 and A.4, which include the entire control group sample placed directly into college-level courses. The status quo pass rates are 72 percent in English (31 percent of 43 percent) and 71 percent in math (10 percent of 14 percent). Compared with the status quo pass rates, the bump-up pass rate is 4 percentage points lower for English and 6 percentage points lower for math. While the bump-up pass rates are slightly lower than the status quo pass rates, students who are bumped up are unlikely to noticeably alter the overall pass rates of the courses they enter (with students who were not bumped up) because they represent a small portion of students in any class.

A similar calculation of fail rates revealed that program group students who were bumped up in English had a fail rate of 19 percent (12 percent of 63 percent), and program group students who were bumped up in math had a fail rate of 13 percent (5 percent of 40 percent).[4] The status quo fail rates are 14 percent in English (6 percent of 43 percent) and 14 percent in math (2 percent of 14 percent). Compared with the status quo fail rates, the bump-up fail rate is 5 percentage points higher for English and 1 percentage point lower for math.

What Happened to Students in the Full Sample?

Appendix Tables A.3 and A.4 show the academic outcomes for all randomized students in English and math, respectively. As noted earlier, the new placement rules did not change course placements for most program group students, as expected. Colleges expected to bump up between 10 and 20 percent of students in each subject from developmental to college level based on the use of multiple measures. About 15 percent of students were referred to gatekeeper English and about 14 percent were referred to gatekeeper math because of multiple measures. They would have been referred to developmental classes under the business-as-usual referral system—as shown by the "Gatekeeper" row under "First semester placement" for each subject.

In the full randomized sample (shown in Appendix Tables A.3 and A.4), placement using MMA caused 4.4 percentage points more students to enroll in a gatekeeper English course and 2.5 percentage points more students to enroll in a gatekeeper math course compared with the control group.

3. If it is assumed that MMA affects outcomes only through its effect on enrollment in college-level courses, and that there are no students who would always defy their placement whether it was made through MMA or business-as-usual methods, then the ratio of the difference in course completion to difference in course enrollment is the complier average causal effect of the intervention. For this completion outcome, it is 55 percent for English and 57 percent for math among those who were induced to take the course by the program. These are impacts of the program among those who received the treatment, whereas the impacts in the tables are among those who were offered the treatment.

4. In English, 64 percent of program group students enrolled in the gatekeeper course, 43 percent passed it, and 12 percent failed it. The rest of the students, 8 percent, withdrew from the gatekeeper course. In math, 40 percent of program students enrolled in the gatekeeper course, 26 percent passed it, and 5 percent failed it. The rest of the students, 9 percent, withdrew from the gatekeeper course.

Slightly more students completed gatekeeper courses in the program group, and more students completed developmental courses in the control group. There was a small positive effect on overall enrollment in the first semester among students who tested for math placement, but the impact disappeared in later semesters.

Effects on Educational Outcomes by Subgroup

The findings presented thus far have been estimates of the overall average effects of multiple measures placement, but there may be different effects for different types of students. It is important to investigate whether MMA placement is equitable for all students and whether there are negative impacts for any subpopulations of students. To better understand if MMA placement is equitable and how it affects achievement gaps, the three confirmatory outcomes (completion of gatekeeper English, completion of gatekeeper math, and college-level credit accumulation) were explored for the following subsets of students:

- Race/ethnicity (Asian, black, Hispanic, white, or other)
- Enrollment status (full time or part time)
- Socioeconomic status (Pell Grant eligible or not)
- High school GPA range (3.0 and higher or below 3.0)
- Status quo developmental placement (one level or two levels below college)
- Eligible for bump-up (eligible in two subjects or one subject)
- Learning and Study Strategies Inventory (LASSI) score (50 and higher or below 50)
- State (Minnesota or Wisconsin)

Impacts on white students were compared with those on students of color to explore effects on race-based achievement gaps. Enrollment status was included to explore if course load is related to students' ability to handle college coursework when given the opportunity to do so. Socioeconomic status was used to explore economic achievement gaps. A GPA grouping near the MMA cutoff was used because if there was a positive effect estimate near the cutoff and no evidence of a higher effect among those with a higher GPA, that may suggest room for a lower cutoff. Status quo developmental placement was investigated because if the intervention works just as well for those placing two developmental courses below (instead of one course below) based on the placement test alone, it could challenge conventional wisdom about those students' remedial needs. Eligibility for bump-up in multiple subjects was included to check for spillover effects for those who were bumped up in multiple courses. The "LASSI" subgroup was added to understand if students with higher LASSI motivation scores (one type of noncognitive assessment) performed better, contributing to a topic of interest in the study of MMA. The team decided to focus on the LASSI, as opposed to the Grit test, because more students in the sample had valid scores on the LASSI. The "state" subgroup was included to account for the different implementation and placement contexts among the four Minnesota colleges and Wisconsin (see Box 3.1).

Also, it is noteworthy that the "enrollment status" and "Pell eligibility" subgroups were defined after random assignment due to the way these data are recorded in the college management information systems (after college enrollment). Because there is a chance that the intervention may have influenced enrollment status, and because Pell information was difficult to collect in some cases, the results of both subgroups should be interpreted with some caution.

Appendix Tables A.5 and A.6 show what proportion of all students who took the placement test were "always developmental," in the bump-up zone, or "always college level" for English and math, respectively, by subgroup. The benefits of MMA are expected only for students in the bump-up zone, so it is important to see which subgroups, if any, fall more into this category. Also, by looking at the

BOX 3.1

Madison College

Madison Area Technical College in Wisconsin employed a different implementation of multiple measures assessment (MMA) than the four colleges in Minnesota. First, the college chose a more personalized referral approach, with less automation than was used in the Minnesota schools: Two faculty advisors attended all advising and registration sessions for incoming students. The advisors looked at a list of students to identify those who were in the program group in the study and potentially eligible to be bumped up to college-level "gatekeeper" classes based on their grade point average (GPA) and Grit test scores. If that was the case, the advisors waved them over to talk further. If students didn't have a Grit score, they were given the short Grit assessment on the spot. The advisors then told them where they placed based on their high school GPA and Grit score. If they were eligible to be bumped up, the advisors followed them over to the registration table and punched in an override code that allowed these students to register for the gatekeeper math or English classes.

However, because these events were voluntary, only about 60 percent of all incoming students attended. As a result, this approach translated into far fewer students being bumped up into college-level courses, and far fewer students who were bumped up enrolling in the college-level course, than the more automated approach used by the Minnesota colleges, in which MMA rules were programmed into the placement testing system.

A "state" subgroup was also run (not shown in the tables) to explore the differential effects of the different implementation and placement contexts among the Wisconsin and four Minnesota colleges. Because so few students were bumped up, there were generally larger impacts on enrollment in and completion of the gatekeeper courses in both subjects among the Minnesota sample compared with the Wisconsin sample. However, these findings should be interpreted cautiously, because they could be attributed to any number of reasons (for example, different implementation processes, different student populations, the use of different noncognitive assessments).

bump-up zone, it may be possible to identify groups of students that tend to have lower placement tests but higher GPAs and/or noncognitive assessment scores.

There were statistically significant differences in the proportion of students across subgroups placed in the bump-up zone.[5] So, even if there were similar impacts across subgroups in the bump-up zone, the subgroups with greater bump-up rates might benefit more. For this reason, the subgroup analyses were performed for the full sample of all randomized students. Appendix Tables A.7 and A.8 show the enrollment into and passing of gatekeeper English. Appendix Tables A.9 and A.10 show the enrollment into and passing of gatekeeper math. Finally, Appendix Table A.11 shows college-level credit accumulation across all subjects.

English Gatekeeper Enrollment and Completion

There were differential impacts on enrollment in the English gatekeeper course among the GPA subgroups, the LASSI subgroups, the status quo developmental placement subgroups, and the two-subject bump-up subgroups. Students in the better-performing subgroups (those with higher GPAs or LASSI scores, higher placement, or placement in both subjects) experienced a bigger impact on enrollment from MMA than those with lower scores. There was a similar pattern for passing the English gatekeeper course, but only among the GPA and bump-up subgroups.

Math Gatekeeper Enrollment and Completion

There were differential impacts on enrollment in and completion of the math gatekeeper course among the GPA subgroups, the LASSI subgroups, the status quo developmental placement subgroups, and the two-subject bump-up subgroups. Students in the better-performing subgroups experienced bigger impacts on enrollment and completion from MMA.

College-Level Credit Accumulation

Among students who were bumped up in both subjects, the program group students earned 4.5 more college-level credits compared with the control group students. When looking at credit accumulation across all subjects, almost all the impact estimates were near zero and not statistically significant, except for students who were bumped up in both math and English. Students who were bumped up in both subjects were referred to college-level courses in both subjects, resulting in more opportunities to take (and earn) college-level credits, which made MMA more effective for them.

Can Cutoff Levels Be Lowered?

Interestingly, there were positive impact estimates on enrollment into English gatekeeper classes among students with lower GPAs, lower LASSI scores, or lower placement levels. There were also positive impact estimates on passing gatekeeper English among students with lower placement levels, and on enrollment into gatekeeper math among students with lower GPAs. For most outcomes, there

5. A chi-square test was used to test the null hypothesis of independence between subgroup and bump-up rate.

are differential impacts for the GPA, LASSI, status quo developmental placement, and two-subject bump-up subgroups, such that students with higher GPA or LASSI scores or with higher placement have higher impact estimates. This is simply a function of the design of the placement system, which is more likely to bump up these students. However, within the bump-up zone (not shown in the tables), these subgroups show no differential impacts. Since impacts within the bump-up zone are not lowered significantly by lower scores on these measures, it is likely that cutoffs for GPA or LASSI might be lowered even further with the expectation of observing positive impacts for students below the current thresholds.

Differences Between English and Math Among Subgroups

Overall, the research team saw slightly different stories for each of the two subjects. There were consistently higher estimated impacts on enrollment in gatekeeper English across all subgroups (Appendix Table A.7) compared with math (Appendix Table A.9). Similarly, there were slightly higher estimated impacts on passing the English gatekeeper course (Appendix Table A.8) compared with the math gatekeeper course (Appendix Table A.10). The differences in the completion impact estimates between the two subjects seem to be driven in large part by higher enrollment in the English course after placement. While these tables present exploratory analyses and show some differences between math and English and some differences for certain groups of students, the results are generally reassuring: All estimated impacts were positive.[6]

6. There are a few subgroups in Appendix Tables A.7 through A.11 with small negative differences on some outcomes, none of which are statistically significant. Within the bump-up zone (not shown in these tables), all subgroups had positive, statistically significant impacts for enrolling in and passing gatekeeper courses.

4

Predictive Utility of Noncognitive Measures

A growing number of colleges use multiple measures assessment (MMA) to determine whether students should be referred to a developmental or college-level course. MMA strategies typically rely on placement test scores and high school grade point average (GPA). But could the placement process be more accurate if other measures were included? This chapter summarizes the findings from an analysis of the predictive utility of noncognitive assessments—the Learning and Study Strategies Inventory (LASSI) motivation scale and the Grit Scale—and considers what this suggests about using such assessments to improve MMA strategies. Understanding the predictive utility of noncognitive measures will help administrators make more informed decisions about incorporating these scales in their MMA systems. The goal of the predictive analysis is to answer the question: How well do the LASSI and Grit noncognitive assessments predict college course completion—alone and in combination with other predictors?

Colleges in this study used the LASSI and Grit noncognitive assessments to bump up students from developmental to college-level classes if they had a noncognitive score above a specific cutoff level. College administrators thought noncognitive assessments would capture something about students' attitudes and behaviors that might help inform how well they will do in college. The predictive analysis in this study assesses how well the LASSI and Grit noncognitive assessments predict success in a college-level course (success is defined as passing with a C or better). Some placement algorithms use these types of predictions to determine placement. They do so by setting a threshold, above which a student is referred to a college-level class and below which they are referred to a developmental class. For example, if a college set the threshold at 60 percent, students with a predicted probability of success of 0.6 or higher would be placed into a college-level class, and students with a predicted probability of success below 0.6 would be placed into a developmental class. Colleges often want to refer students with a high probability of success directly to college-level courses, so those students don't spend time and resources on a developmental course they might not need. For students with a lower probability of success, colleges prefer to place them in developmental courses, in the hope that this will improve their long-term likelihood of succeeding in the college-level course.

Considerable thought went into the selection of the LASSI and Grit tests by the colleges. See Box 1.1 in Chapter 1 for more information about why these noncognitive assessments were selected by the colleges.

The findings and takeaways from this analysis are discussed in this chapter. Most importantly, this analysis found that the LASSI motivation scale and Grit noncognitive assessments did not improve the predictive accuracy beyond that of high school GPA for placement in either English or math.

However, it is important to note that this analysis only included two noncognitive assessments, so there is a limit on what can be inferred about noncognitive assessments as a whole. There are many other noncognitive measures that may have more predictive utility than the two used in the current analysis.

The Empirical Approach

The predictive analysis uses data from the colleges to develop models that can quantify the potential contribution of noncognitive assessments to placement accuracy. The primary courses of focus are English and math for those students who took placement tests for these subjects. This analysis relies only on observable information about success in college-level courses by restricting the sample to students who took a college-level course in their first semester after placement.[1] The sample is further limited to students with scores on the tests for a given subject.[2]

Several predictive models are run to predict each student's likelihood of succeeding in college-level English or math. Table 4.1 provides an example of the empirical approach. First, students are classified

TABLE 4.1 Placement Recommendations Versus Success in College-Level Course

PLACEMENT	SUCCEEDED	DID NOT SUCCEED
College level	(a) Correct placement (true positive)	(b) Incorrect placement (false positive)
Developmental level	(c) Incorrect placement (false negative)	(d) Correct placement (true negative)

(in the rows) by their placement recommendation—they can either be referred directly to a college-level course or referred to a developmental course. Referral may be based on a single measure or on multiple measures combined in various ways during modeling. Each model produces an estimated likelihood of success, on a scale from 0 to 1, for each student. For a given threshold, students with estimated likelihoods above the threshold are recommended to be placed in college-level courses, and students with estimated likelihoods below the threshold are not recommended for college-level courses. Different placement decisions are estimated in the rows of the table by specifying different thresholds. Then, students are classified (in the columns) by their observed success in the college-

1. Students with developmental placements or corequisite placements were considered to have "non-college-level" placement recommendations, so their performance in college-level courses was ignored. Their performance in the college-level course is likely affected by the prerequisite, which occurs after the placement process is over, and is correlated with placement measures that are used to predict success in college-level courses.

2. Of all the English placement tests, only the reading comprehension test was assessed, because too few colleges and/or students had data for the sentence skills or Writeplacer tests. One of the Minnesota colleges was excluded from the math analyses because not enough students enrolled in college-level math courses at this site. The arithmetic math test was not assessed because fewer than 500 students took this test across all sites, with most students coming from one or two sites.

level course they took. The performance of a model can be assessed by comparing predictions of success to students' observed outcomes.

A placement is considered correct if a student is placed in the college-level course and passes the course (cell a), or if a student is not placed into the college-level course and would not have passed the course had they taken it (cell d). Otherwise, there are two types of incorrect placements: A student who is placed in the college-level course who does not pass is "overplaced" (cell b),[3] and a student who is not placed in the college-level course but would have passed had they taken the course is considered "underplaced" (cell c). The challenge in filling in this table is that cells c and d are usually not observed. That is, students who were referred to developmental courses and did not immediately take a college-level course do not have a college-level course outcome. This analysis restricts the sample to students who took a college-level course in their first semester, ensuring an outcome for all students.

For each threshold, a new 2 x 2 table can be filled, and various performance metrics can be computed:

The true positive rate: Among students who would succeed in a college-level course, this is the proportion correctly placed (in the table, *a / (a+c)*).

The false positive rate: Among students who would not succeed in a college-level course, this is the proportion incorrectly placed (in the table, *b / (b+d)*).

The predictive accuracy: This is the proportion of all students who were correctly placed (in the table, *(a+d) / (a+b+c+d)*).

The Modeling Approach

A logistic regression was used to model the relationship between various measures and college-level success. Two machine learning algorithms—LASSO (Least Absolute Shrinkage and Selection Operator) and Random Forest—were also used to try to improve predictive performance. However, neither of these machine learning approaches provided improvements in the predictive performance—perhaps because there were only a few measures in each model, or because there were no meaningful interaction effects, or because the nature of the relationships between the predictors and outcomes is generally linear. Therefore, the figures in Appendix B focus on logistic regression; explanations for how to read these figures can be found in Appendix A. Findings from all models are in Appendix C.[4]

3. The term "overplaced" has been used in previous MMA research (such as Scott-Clayton, 2012). However, this terminology may be misleading at times. For example, imagine students who placed in college-level English but would have failed this course regardless of their placement (even if they took the developmental English course first). In this scenario, the students were not "overplaced"—they would have failed no matter what.

4. Random Forest algorithms call for several different specifications (or tuning parameter settings), so each predictor set was used several times with a variety of settings in an attempt to capture the best possible specifications. However, the tables in the appendix include only one Random Forest model for each predictor set. This was done to simplify the contents of the tables. The model with the highest AUC ROC was chosen.

A useful empirical summary of a model's performance, across all potential thresholds, is the area under the curve (AUC) of a receiver operator curve (ROC). The ROC shows the trade-offs between the true positive rate and the false positive rate at each possible threshold. Predictive models equivalent to a random coin flip (roughly 50-50) have an AUC ROC = 0.50, while those that are 100 percent correct have an AUC ROC = 1 (the higher the AUC, the better).

Other summaries of a model's predictive performance could be examined at specific thresholds. For example, the model's accuracy can be calculated using a threshold of 50 percent, and again using a threshold of 60 percent, and so on. Table 4.2 (discussed in the next section) compares the models' accuracies at various thresholds.[5] The accuracy is also compared with a naïve model where all students were placed into college-level courses.

TABLE 4.2 Predictive Accuracy of All Models Compared with All Students Being Placed Directly into College-Level Courses

MODEL	READING COMPREHENSION SAMPLE	ELEMENTARY ALGEBRA SAMPLE	COLLEGE-LEVEL MATH (CLM) SAMPLE	ELEMENTARY ALGEBRA AND CLM SAMPLE
Test only	0.62	0.59	0.64	0.64
Noncognitive only	0.64	0.62	0.66	0.65
GPA only	0.68	0.64	0.71	0.71
Test + noncognitive	0.62	0.59	0.63	0.64
Test + GPA	0.67	0.63	0.70	0.70
Test + GPA + noncognitive	0.67	0.63	0.69	0.68
Naïve model (all placed in college level)	0.69	0.68	0.71	0.71

SOURCE: Placement data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

Findings

- **Across both math and English contexts, the LASSI and Grit noncognitive assessments did not do a good job of predicting success in college-level courses when used alone.**

Table 4.2 shows the accuracy of models that used various measures, alone and in combination, to predict success in college-level courses. Models that used only these noncognitive assessments were less accurate compared with models with GPA. However, the models with these noncognitive assessments were slightly more accurate compared with ACCUPLACER alone, but the improvements

5. Each model's predictions were classified using a threshold that mimicked what colleges did in reality: It mimicked the observed proportion of students placed into college-level courses (among the students in each model's sample). These thresholds ranged from 0.55 to 0.68, which are moderate thresholds that most colleges would probably use when making placement decisions.

in predictive performance were minimal. While GPA seems to be the best predictor relative to these other measures, none of these models had great predictive performance overall. The highest AUC of any model was only 0.66, which is closer to a coin flip than perfect predictive performance.

A similar pattern emerged when summarizing predictive performance across all thresholds. The figures in Appendix B show the ROCs that summarize predictive performance across all possible thresholds. These figures consistently show that the models with GPA outperform the noncognitive models and the ACCUPLACER models—though the difference in predictive performance between GPA and other measures is smaller in the math context (meaning the lines in the figures are closer together). This drop in predictive performance is evident when comparing Appendix Figure B.1 to Appendix Figure B.2. The GPA model has more area under the curve compared with the other models in the English context (Appendix Figure B.1), but in the math context (Appendix Figure B.2), the elementary algebra model has almost as much area under the curve as the GPA model. This suggests two things: (1) There are differences between English and math in terms of how well a model can predict success in college-level courses, and it appears success in college-level English may be easier to predict, and (2) GPA seems to be a better predictor of college-level success in either subject, but noncognitive assessments appear to hold little predictive utility beyond ACCUPLACER alone.

- **Across both math and English contexts, none of the predictive models outperformed the scenario where all students were placed into college-level courses, but GPA performed similarly well.**

In Table 4.2, the model with the highest accuracy across all contexts was the naïve model that allowed all students to take college-level courses. However, the models with GPA had a similar accuracy (it was almost identical in most of the samples). Given that these accuracies were only calculated for a handful of thresholds, it is possible the other models would outperform the naïve model if a different threshold were used. It is also worth noting that other metrics of predictive performance may have shown a different pattern, because the naïve model places more students into college-level courses than other placement models. If all students can take college-level courses, more of them may end up failing or withdrawing because they were not ready for the college level. Thus, using predictive models can minimize false positive rates, because the students who are not ready for college-level courses will be placed into developmental courses first. On the other hand, research has shown that placing students into developmental courses increases their chances of dropping out and not moving on to college-level courses, potentially increasing false negative rates. So, placing all students into college-level courses may have some benefits compared with more restrictive models that only place a few students into college-level courses.

- **The noncognitive assessments did not improve predictive performance among older students.**

An additional exploratory analysis was performed (not shown in the figures or tables) to see if the noncognitive assessments have more predictive utility for older students. This analysis was done by comparing students aged 25 years or older with students younger than 25.[6] When predicting success in college-level English, the GPA model was most predictive for the younger students, but all mod-

6. There were fewer older students, so not all the assessments were analyzed for this additional exploratory analysis. Only the reading comprehension models were run to predict success in college-level English.

els performed poorly for older students. However, the sample size was small for the older group of students (N = 530), so these results should be interpreted with caution. Also, only two noncognitive assessments were used in these analyses, so colleges should continue to explore other noncognitive assessments to help improve predictive modeling, especially when high school GPA is not available.

Limitations

The predictive analyses relied only on students who enrolled in college-level courses immediately after placement, so the generalizability of the findings is limited. The findings do not include information about the extent to which students who did *not* enroll in a college-level course *would have* succeeded if they had taken the college-level course. Yet these students are part of the target population of the placement assessment process and ideally would be included in these analyses. The underlying assumption of this analysis is that the relationship between the predictors and the outcomes would be the same for students going into developmental courses, but this is a strong assumption, which may not be true.

Also, students who enrolled in college-level courses were more likely to have higher test scores and higher GPAs than those who did not enroll. Moreover, the range of scores on the placement tests is more restricted than the range of GPAs, because only students who enrolled in college-level courses were included in this analysis, and it is unlikely this sample included anyone with very low ACCUPLACER test scores, unless those students went against their placement recommendation and enrolled in college-level courses instead of developmental courses. This means that the range of scores included in these analyses is not representative—again leading to limited generalizability and likely an understatement of the predictive utility of the placement test scores.

Relying on students who enrolled in college-level courses also makes cells c and d in Table 4.1 dependent on students who do not follow their placement decisions (students who were placed below the college level but decided to enroll in college-level courses anyway). Because of this, it is possible the true positive rates and false positive rates are overstated, so the performance metrics should be interpreted with some caution.

Finally, it is worth noting that this predictive analysis focuses on outcomes, not impacts. The models are predicting success in college-level courses, and not the impact of MMA placement on success in college-level courses, the latter being of more interest in most MMA systems.

5

Cost

Setting up and administering multiple measures assessment (MMA) took substantial effort by college staff, even after the initial decisions about high school grade point average (GPA) and noncognitive assessment cutoff scores had been made. This effort included reprogramming placement system platforms and registration systems to recognize high school GPAs for placement into college-level courses despite ACCUPLACER results below the usual thresholds; changing data collection, entry, and student communications at the admissions stage; changing the advising process to include an explanation of multiple measures; and administrative planning and oversight of these changes. These activities were additional to the business-as-usual testing and placement processes, and each college approached these additional activities in slightly different ways depending on the baseline procedures and their specific MMA criteria. Cost data were captured at the end of each semester using staff questionnaires on hours spent on these activities that would not have occurred without the MMA intervention. Using these data, the team found that this implementation effort cost the colleges about $33 per student who went through the placement process during the three semesters of the study. This cost is comparable to those of other programs that focus on behavioral nudges, such as the EASE informational campaign ($16), with comparable cost-effectiveness (per credit earned) as well.[1]

Table 5.1 breaks down the direct costs of the MMA programs, which include administration, staffing, and materials. These costs represent effort and materials that would not have been incurred under business-as-usual placement, and were collected from participating colleges, reporting additional hours spent by staff members in the categories presented and their corresponding wage or salaries. For materials, the number of Learning and Study Strategies Inventory (LASSI) tests administered were multiplied by the $3.50 per-test cost. The per-student averages include all students, regardless of enrollment status, not just those in the bump-up zone. This is because the placement system itself was scaled to all students, regardless of whether or not they were bumped up. Furthermore, per-student costs include both program and control students in the denominator. This is because once the system was set up, program group placement rules could have been applied to control students at no additional cost whatsoever (the differing placement results were a contrivance necessary for the randomized controlled trial). Excluding control students when dividing the direct cost by the number of students would erroneously double the per-student costs a college should expect for implementing such a placement system.

1. Anzelone, Weiss, Headlam, and Alemañy (2020). Amount adjusted to 2021 dollars. MDRC's Encouraging Additional Summer Enrollment (EASE) study used behavioral insights and a financial incentive with the goal of boosting enrollment rates.

TABLE 5.1 Direct Cost of the Program per Sample Member

PROGRAM COMPONENT	PER HOUR ($)	PER COLLEGE SEMESTER RANGE ($)	PER STUDENT ($)	PERCENTAGE OF TOTAL (%)
Personnel				
Information technology	46	302 - 6,830	1	2
Admissions/testing/advising	35	20,872 - 112,607	18	54
Faculty and registrar	52	955 - 36,036	4	12
Administrative staff	57	4,835 - 75,872	7	20
Technical assistance	84	1,685 - 3,370	1	2
Materials				
Noncognitive assessments		848 - 6,076	3	9
Total direct cost		59,747 - 211,430	33	

SOURCE: MDRC calculations based on program expenditure data from the four Minnesota colleges and one Wisconsin college.

NOTES: Rounding may cause slight discrepancies in sums and differences.
Program costs are based on direct costs during the first three semesters of the program. Per-hour amounts are averages of hourly rates by category including overhead and benefits provided by the colleges. The discount rate used for program costs is 3 percent. All costs are shown in constant 2021 dollars.

Direct Cost of MMA

The total direct cost per program group member is $33 (a total that includes program group members who did not enroll in college). About 54 percent of the direct cost of the program, $18 per student, comes from those people guiding students through the process that leads to course registration: admissions, testing, and advising staff. Administrative staff, who managed the program, represented about 20 percent of the direct cost, $7 per program member. Faculty and registrars represented about 12 percent of the direct cost, $4 per program member. The smallest components, information technology and technical assistance,[2] make up about 4 percent of the cost. Noncognitive assessments make up 9 percent of the programs' direct cost, at $3 per student.

Sensitivity Analyses

Per-student direct costs varied from $20 to $37 across the five participating colleges. Sensitivity analyses around necessary technical assistance time assumptions (from 20 to 40 hours per semester per college) have a negligible effect on the direct cost estimate.

2. MDRC staff provided technical assistance to the participating colleges. MDRC staff members spent a total of approximately 30 hours on technical assistance that was not associated with the research study evaluation per college once the study had begun. Using national salary and benefit averages for the category of Education Administrators—Management and Technical Consulting from the Bureau of Labor Statistics, this cost was approximately $2,340 per college, less than a dollar per student.

These direct costs do not include the planning and start-up costs associated with the pilot phase of the program, which preceded the randomized controlled trial. The direct costs per college of that initial phase were reported in the interim report as $52,402 per college.[3] If these start-up costs were added to the costs incurred during the scaled implementation, the per-student cost would rise to $48 over the period of the study. If the direct costs shown in Table 5.1 were measured per *enrolled* student, they would be $40, about 21 percent higher than the dollar amounts identified above because many students did not enroll in college.

Finally, the cost estimates presented here are for implementing MMA for both math and English. Costs could be as little as half these amounts if MMA were implemented for only one subject, but effects would be limited correspondingly as well.

Indirect and Net Costs

Table 5.2 adds the indirect cost and revenue to the colleges, brought about by the impacts of the program, to calculate net cost, as well as the incremental cost-effectiveness. Indirect costs are estimated based on the average number of additional credits attempted by the program students compared with the control group students. The impact of MMA on all credits attempted when developmental and college-level courses are included is quite modest, and not statistically significant. For the most part, program group students substituted college-level course-taking for developmental course-taking. This analysis averages two approaches. A lower-bound estimate assumes that the indirect costs equal zero—that is, that the college incurs no additional costs when more students enroll and/or when students attempt additional credits. An upper-bound estimate is based on average instructional costs per credit from the Integrated Postsecondary Education Data System.

It is unlikely that every additional credit attempted by a student costs the college as much as the average credit attempted, and it is also unlikely that there is zero cost to the college for additional credits attempted. An average of these two estimates—the midpoint between the upper and lower bounds—is therefore used as the primary estimate of indirect costs: $30 per program group student. This amount is almost exactly offset by the expected tuition revenue associated with additional credits attempted: $30 per program group student.

From the student perspective, developmental courses might be considered additional costs, and reducing them, a savings. With an average of a half-credit reduction in developmental credits attempted across the full sample, program group students saved about $100 in tuition that would have been spent on developmental courses, on average.

The net cost to the colleges is presented in the second section of Table 5.2. The net cost is calculated by adding the direct cost to the indirect cost and subtracting state funding and tuition revenue. The net cost is defined as the difference between the total program group cost and the total control group cost. The net cost is approximately the same as the direct cost, $33 per program group member.[4]

3. Cullinan et al. (2019). Amount adjusted to 2021 dollars.

4. Societal net cost would not subtract tuition revenue, which is a series of transfer payments, not negative costs. Without this adjustment, the net cost would total to $52 per student instead.

TABLE 5.2 Net Cost per Sample Member and Cost-Effectiveness Values from the College Perspective

OUTCOME	DIFFERENCE (IMPACT)	90% CONFIDENCE INTERVAL	
Direct cost: cost of primary program components ($)	33		
Indirect cost: cost of additional credits attempted due to the program ($)	30	-49	114
Indirect revenue: tuition from additional credits attempted due to the program ($)*	30	-50	116
Net cost per group member ($)	33		
Total college credits earned	0.2	-0.1	0.5
Incremental cost per additional credit earned ($)	136	N/A	60
Completed math gatekeeper course (percentage points)	0.9	0.1	1.6
Incremental cost per additional completed math course ($)	3,620	32,578	2,036
Completed English gatekeeper course (percentage points)	1.7	0.6	2.8
Incremental cost per additional completed English course ($)	1,916	5,430	1,164
Sample size (total = 17,203)			

SOURCE: MDRC calculations from program-specific expenditure data, transcript data, and financial and enrollment data from the Integrated Postsecondary Education Data System.

NOTES: Rounding may cause slight discrepancies in sums and differences. All dollar values have been rounded to the nearest whole dollar.

Tests of statistical significance have only been performed on outcome measures, not costs. All outcomes are cumulative over three years. All costs are shown in constant 2021 dollars.

*This revenue represents transfers to the college from students and government via tuition, financial aid, and scholarships. It is not a societal cost offset or benefit, but is included here to represent the college perspective.

Cost-Effectiveness

A cost-effectiveness analysis expresses the cost of interventions as the cost per unit of a desired outcome, for example, the additional cost per additional credit earned. The incremental cost per additional outcome caused by the program can be compared with that of other programs. This ratio might be useful when comparing programs with similar impacts on the same outcome, giving policymakers more than one estimate of the cost of achieving those impacts. This cost-effectiveness analysis considers the cost per college credit earned as the primary outcome because of the available outcomes, it best represents the overall gains in human capital from implementing MMA. Costs per gateway English and math course within three semesters are secondary outcomes for this cost-effectiveness analysis. These estimates spread costs across all students who were offered MMA, including those who enrolled less than full time or dropped out.

The bottom half of Table 5.2 shows the cost-effectiveness calculations for the program from the perspective of the college. The first row below the net cost shows the average impact on credits earned in three semesters for the entire study sample.[5] The incremental cost-effectiveness ratio is $136 per additional credit earned. This is about twice the $69 cost per additional credit earned of the EASE informational campaign, another low-cost behavioral-style intervention.[6] Because passing gatekeeper courses is also an important outcome for MMA systems, cost-effectiveness is calculated for those outcomes as well. As shown in the table, the incremental cost per additional gatekeeper course passed was $3,620 for math and $1,916 for English.

Another comparable program with a cost-effectiveness analysis is found in the report on the MMA systems implemented by seven State University of New York (SUNY) colleges noted earlier.[7] The per-student direct cost of the MMA systems at the five colleges in Minnesota and Wisconsin was approximately one-fifth that of the MMA systems at the seven SUNY colleges over a similar time period ($158). But because the reduction in course-taking was more modest in Minnesota and Wisconsin, direct costs were not offset by indirect (negative) costs, as was the case in the SUNY trial. This means that from a societal perspective, the Minnesota and Wisconsin version of MMA is less cost-effective per additional credit earned. It is possible that had the SUNY cost analysis considered tuition revenue (from the college perspective, as this analysis does), the net cost to the college would have been higher than the net (societal) cost presented in that report because of the loss of tuition from developmental courses.

The cost-effectiveness ratios in Table 5.2 suggest that in the early semesters of new MMA system implementation, such as the one implemented for this study, costs per student outcome are considerable, and should be weighed carefully against other options. However, the MMA system implemented by the one Wisconsin and four Minnesota colleges sought to incorporate high school GPA into an existing placement test platform. If high school GPA could be used in an automated way without needing to recode placement test systems, such placement could possibly achieve similar results for a much lower cost. Likewise, the proportion of the sample whose placement was changed because of MMA was below 16 percent, while between one-third (English) and two-thirds (math) of placed students remained in developmental placement under either system. If a significant number of these were students who were placed differently under MMA, and if similar impacts were observed for them, the cost per college credit earned and per gatekeeper course passed would be significantly lower. Finally, the placement system had the highest costs the first semester. If used for additional semesters beyond the third, per-student costs would continue to fall because the higher costs of the first semester would be spread over a longer time period.

5. These impact estimates differ slightly from those shown in Appendix Tables A.3 and A.4 because this sample combines all students who tested in either subject.

6. Anzelone, Weiss, Headlam, and Alemañy (2020). Amount adjusted to 2021 dollars.

7. Barnett, Kopko, Cullinan, and Belfield (2020). Amount adjusted to 2021 dollars.

6

Conclusions

The first-semester impacts presented in the previous report showed that the use of multiple measures assessment (MMA) accomplished its first-semester goal of changing students' placements when the grade point average (GPA) and noncognitive cutoffs were met and its goal of increasing enrollment into college-level courses. The three-semester impacts in the current study confirm the early findings on enrollment and further suggest that MMA placement has a positive impact on academic outcomes, and importantly, the impacts appear to be robust across all student subgroups.

Program group students in the bump-up zone who were placed into college-level English were 16 percentage points more likely to have completed the gatekeeper English course by the end of their third semester than their control group counterparts.

Program group students in the bump-up zone who were placed into college-level math were 11 percentage points more likely to have completed the gatekeeper math course by the end of their third college semester than their control group counterparts.

Overall, all subgroups of students benefited from multiple measures placement, and MMA generally had positive impact estimates on enrollment in and completion of gatekeeper courses in English and math.

The predictive analysis found that GPA was the best of the available predictors of success in college-level courses. The Learning and Study Strategies Inventory (LASSI) and Grit noncognitive assessments appeared to add no predictive value above and beyond that of GPA.

Implementing MMA cost the colleges $33 per student over the business-as-usual placement process. It is comparable to, but somewhat costlier than, the per-student and per-credit-earned costs of the EASE informational campaign. The MMA cost could likely be lowered over time either through continued use or by tweaks to the implementation.

These findings show that MMA increased gatekeeper course completion when students who met certain high school GPA or noncognitive assessment thresholds but who didn't meet the usual ACCUPLACER thresholds were bumped up into college-level courses instead of developmental prerequisites. These MMA systems worked well for every subgroup of students, across race/ethnicity, gender, age, and Pell status, and worked well for both math and English. They also worked well for bumped-up students with lower levels of preparation as observed in their multiple measures scores. This suggests that colleges can use such systems with confidence that students from all these subgroups will benefit on average.

Room for Improvement

The success of MMA placement systems is highly dependent on whether students enroll in the course they are placed into. For example, in English, more program group students enrolled in the English gatekeeper course compared with control group students, and more program group students completed the English gatekeeper course. In math, the impacts on the completion of the gatekeeper course were lower. Because far fewer students who placed into college-level math ended up enrolling in the math gatekeeper course, impacts on the completion of gatekeeper math were much lower compared with English.[1] This suggests that colleges implementing MMA should focus not only on delivering the placement result, but also on encouraging students to enroll in college-level math and English their first semester. Other research has shown that it is possible to increase enrollment with the right messaging—for example, the EASE study saw positive impacts on summer enrollment when using informative messaging.[2] Adding messaging that encourages enrollment into gatekeeper courses might get more students to enroll in college-level courses in English and math, leading to more completion of college-level courses in both subjects.

Multiple measures placement systems may improve outcomes for more students by lowering GPA cutoffs. Among all randomized students, students with GPAs below 3.0 experienced positive impacts on enrollment into English and math gatekeeper courses, and among students in the bump-up zone, the impacts on enrollment in and passing of these courses were not lowered by lower GPAs. These findings suggest that lowering the GPA cutoffs further might increase the enrollment of additional college-ready students into the gatekeeper courses, thereby increasing their completion.

Future Research

The current study only investigated two noncognitive assessments—the LASSI motivation scale and the Grit Scale—and found that these two assessments may not have any additional predictive utility beyond that of GPA or ACCUPLACER when predicting college-level success. However, these two noncognitive assessments do not represent all noncognitive measures. Future research should investigate the predictive utility of other noncognitive measures to better understand how such measures in general can improve MMA placement. Furthermore, there are many more common uses for noncognitive assessments, such as identifying additional supports for individual students based on their responses, that this study does not address.

The current study assessed the effectiveness of a simpler MMA placement system compared with other studies (for example, the State University of New York study mentioned in Chapter 1). Given

1. Gatekeeper completion rates divided by enrollment rates are almost the same for math and English—65 percent and 68 percent, respectively—suggesting that most of the difference in impacts on course completion is attributable to differences in enrollment rates.

2. Anzelone, Weiss, and Headlam (2020).

the generally positive findings seen in the current study, a simpler MMA approach may be as beneficial as a more complicated one, as well as being less costly. However, under the simpler MMA approach, fewer students' placement was changed. This is not a function of the system's complexity, but an exogenous choice of cut-off thresholds made by faculty and administrators. MMA systems can change the placement of more students without being overly complicated in order to ease implementation burdens.

The current study only calculated impacts during three semesters after initial placement, but the longer-term effects of MMA placement are still not well understood. Future research should examine the effects of MMA placement on students' academic outcomes beyond three semesters. An upcoming study will do just that. Funded by Ascendium Education Group, this study will collect graduation data on students for three years after placement, allowing researchers to see how MMA affects students' long-term academic outcomes.

APPENDIX

A

Supplemental Tables

APPENDIX TABLE A.1 Baseline Characteristics of the Full Sample

CHARACTERISTIC (%)	PROGRAM GROUP	CONTROL GROUP	BOTH GROUPS
Age			
20 and under	57.7	56.7	57.3
21-30	20.8	20.7	20.7
31 and over	7.9	8.4	8.1
Age missing	13.6	14.2	13.8
Gender			
Male	37.3	37.7	37.4
Female	49.1	48.1	48.7
Gender missing	13.6	14.2	13.8
Race/ethnicity			
Asian	7.5	7.1	7.4
Black	14.1	14.1	14.1
Hispanic	10.9	11.0	10.9
White	47.3	46.8	47.1
Other	6.2	6.2	6.2
Race/ethnicity missing	14.1	14.7	14.3
Enrollment status			
Full time	42.5	41.1	41.9
Part time	34.5	34.0	34.3
Enrollment status missing	23.1	24.9	23.8
Pell eligibility			
Yes	32.1	31.8	32.0
No	44.8	45.1	44.9
Pell eligibility missing	23.0	23.2	23.1
Sample size	10,476	6,727	17,203

SOURCE: Demographic data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Distributions may not add to 100 percent because of rounding.

Enrollment status represents enrollment in the first semester. For one of the sites, this was determined based on credits attempted in the transcript data.

APPENDIX TABLE A.2 Multiple Measures Assessment Scores Among the Full Sample of Randomized Students

TEST	PROGRAM GROUP	SD	CONTROL GROUP	SD	DIFFERENCE	P-VALUE
ACCUPLACER Scores[a]						
Arithmetic	50.4	26.6	50.7	26.0	-0.3	0.671
Elementary algebra	60.2	27.6	59.9	26.5	0.3	0.532
College-level math	39.8	18.0	39.6	18.0	0.2	0.687
Reading comprehension	77.2	21.5	78.1	21.1	-1.0 **	0.022
Sentence skills[b]	77.9	20.8	79.7	19.9	-1.7 ***	0.008
High school GPA (%)						0.764
3.5-4.0	9.7		9.5			
3.0-3.4	14.4		14.7			
2.5-2.9	17.3		16.8			
2.0-2.4	10.8		11.1			
1.9 or lower	6.2		5.9			
GPA missing	41.6		42.0			
LASSI score (%)						0.918
50-100	34.3		34.3			
0-49	23.9		24.1			
LASSI missing	41.8		41.5			
Grit score	3.7	0.5	3.8	0.7	-0.1 **	0.022
Sample size (total = 17,203)	10,476		6,727			

SOURCE: Test scores, high school GPA, and LASSI and Grit scores provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.

Statistical significance levels are indicated as: *** = 1 percent, ** = 5 percent, * = 10 percent.

The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

To assess differences between the research groups, chi-square tests were used for categorical variables and two-tailed t-tests were used for continuous variables.

SD = standard deviation, GPA = grade point average, LASSI = Learning and Study Strategies Inventory.

[a]ACCUPLACER test scores can range from 0 to 120.

[b]Only Normandale Community College used the sentence skills test to determine course placement for English. The other three Minnesota colleges used the reading comprehension test, and Madison used a combination of the two tests to determine course placement for English.

APPENDIX TABLE A.3 Academic Outcomes After Three Semesters Among All Randomized Students Who Tested for English

OUTCOME	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL LOWER BOUND	90% CONFIDENCE INTERVAL UPPER BOUND	P-VALUE
First-semester placement						
Gatekeeper (%)	63.9	49.1	14.8	13.7	15.8	0.000
Developmental (%)	36.1	50.9	-14.8	-15.8	-13.7	0.000
Three-semester outcomes						
Gatekeeper (%)						
Enrolled	47.6	43.1	4.4	3.1	5.8	0.000
Completed (C or higher)	33.3	31.2	2.1	0.8	3.4	0.009
Failed	7.3	6.0	1.3	0.6	2.1	0.004
Withdrew	7.2	6.1	1.1	0.3	1.8	0.020
Developmental (%)						
Enrolled	16.6	21.0	-4.4	-5.4	-3.4	0.000
Completed (C or higher)	11.1	15.2	-4.1	-5.0	-3.2	0.000
Failed	3.2	3.9	-0.7	-1.3	-0.2	0.024
Developmental	2.9	2.7	0.2	-0.3	0.7	0.440
College level						
Credits earned (C or higher)	2.03	1.98	0.05	-0.03	0.12	0.284
Number of courses completed	0.62	0.61	0.01	-0.01	0.04	0.334
All subjects						
Enrolled during first semester (%)	78.9	77.8	1.1	0.2	2.0	0.054
Enrolled during second semester (%)	63.1	63.6	-0.5	-1.8	0.9	0.581
Enrolled during third semester (%)	45.4	45.3	0.0	-1.4	1.5	0.962
Number of semesters enrolled	1.87	1.87	0.01	-0.02	0.04	0.715
Total credits attempted	20.65	20.62	0.03	-0.35	0.41	0.899
Total credits earned	15.08	15.29	-0.21	-0.60	0.19	0.384
College-level credits earned (C or higher)	12.48	12.27	0.21	-0.15	0.57	0.329
Developmental credits earned	1.69	2.13	-0.44	-0.54	-0.33	0.000
College-level courses completed	4.26	4.22	0.04	-0.08	0.16	0.605
Sample size (total = 12,046)	7,405	4,641				

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.

The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

APPENDIX TABLE A.4 Academic Outcomes After Three Semesters Among All Randomized Students Who Tested for Math

OUTCOME	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL LOWER BOUND	90% CONFIDENCE INTERVAL UPPER BOUND	P-VALUE
First-semester placement						
Gatekeeper (%)	22.5	9.6	12.9	12.1	13.7	0.000
Developmental (%)	72.6	86.7	-14.0	-14.8	-13.3	0.000
Three-semester outcomes						
Gatekeeper (%)						
Enrolled	16.7	14.2	2.5	1.5	3.4	0.000
Completed (C or higher)	11.1	10.1	0.9	0.1	1.7	0.058
Failed	2.2	1.6	0.6	0.2	1.0	0.007
Withdrew	3.2	2.3	0.8	0.4	1.3	0.002
Developmental (%)						
Enrolled	23.4	26.8	-3.4	-4.5	-2.3	0.000
Completed (C or higher)	14.8	18.6	-3.9	-4.8	-2.9	0.000
Failed	6.3	6.4	-0.1	-0.7	0.6	0.880
Withdrew	4.7	4.5	0.2	-0.4	0.8	0.568
College level						
Credits earned (C or higher)	1.24	1.17	0.07	0.01	0.13	0.060
Number of courses completed	0.36	0.33	0.02	0.00	0.04	0.033
All subjects						
Enrolled during first semester (%)	78.9	77.7	1.2	0.4	2.1	0.017
Enrolled during second semester (%)	62.9	63.5	-0.6	-1.8	0.6	0.426
Enrolled during third semester (%)	45.5	45.4	0.2	-1.2	1.5	0.847
Number of semesters enrolled	1.87	1.87	0.01	-0.02	0.03	0.625
Total credits attempted	20.49	20.56	-0.07	-0.42	0.27	0.729
Total credits earned	15.20	15.41	-0.20	-0.56	0.15	0.347
College-level credits earned (C or higher)	12.76	12.51	0.25	-0.07	0.57	0.202
Developmental credits earned	1.54	1.99	-0.45	-0.54	-0.36	0.000
College-level courses completed	4.32	4.27	0.05	-0.06	0.16	0.432
Sample size (total = 15,002)	9,106	5,896				

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.
The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

APPENDIX TABLE A.5 Multiple Measures English Placement by Subgroup

	ALWAYS DEVELOPMENTAL			BUMP-UP ZONE			ALWAYS COLLEGE LEVEL		
OUTCOME	%	N	P-VALUE	%	N	P-VALUE	%	N	P-VALUE
Race/ethnicity			0.000			0.000			0.000
Asian	45.5	375		17.7	146		36.8	304	
Black	53.6	922		19.5	336		26.9	462	
Hispanic	43.7	575		16.6	219		39.7	522	
White	24.5	1,403		13.5	773		62.0	3,543	
Other	31.7	236		16.6	124		51.7	385	
			0.000			0.190			0.000
Students of color	45.8	2,109		17.9	826		36.3	1,672	
White	24.5	1,403		13.5	773		62.0	3,543	
Enrollment			0.120			0.000			0.000
Full time	29.7	1,525		16.9	869		53.5	2,749	
Part time	39.0	1,612		14.1	582		46.9	1,941	
Pell eligible			0.001			0.300			0.000
Yes	41.7	1,662		17.9	713		40.5	1,614	
No	27.7	1,481		14.1	753		58.1	3,104	
GPA range			0.000			0.000			0.615
3.0 or higher	8.2	237		29.4	846		62.4	1,796	
Below 3.0	43.4	1,870		15.5	669		41.0	1,766	
LASSI range			0.002			0.000			0.000
50-100	27.8	1,144		24.4	1,002		47.8	1,967	
0-49	44.0	1,297		12.0	354		43.9	1,294	
Status quo English placement			0.000			0.000			
1 level below college level	59.3	2,128		40.7	1,462				
2 levels below college level	84.9	1,887		15.1	336				
In the bump-up zone						0.000			
For both math and English				100.0	488				
For either math or English	4.4	98		59.5	1,335		36.1	811	
Sample size		4,391			1,814			5,841	

SOURCE: Demographic data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Distributions may not add to 100 percent because of rounding.

The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

To assess differences between the research groups, chi-square tests were used for categorical variables and two-tailed t-tests were used for continuous variables.

Enrollment status represents enrollment in the first semester. For one of the sites, this was determined based on credits attempted in the transcript data.

APPENDIX TABLE A.6 Multiple Measures Math Placement by Subgroup

	ALWAYS DEVELOPMENTAL			BUMP-UP ZONE			ALWAYS COLLEGE LEVEL		
OUTCOME	%	N	P-VALUE	%	N	P-VALUE	%	N	P-VALUE
Race/ethnicity			0.000			0.000			0.000
Asian	62.6	671		17.1	183		20.3	218	
Black	74.4	1,585		14.5	309		11.1	237	
Hispanic	78.4	1,323		11.0	185		10.6	179	
White	71.6	5,233		14.5	1,061		13.9	1,017	
Other	73.4	689		11.0	103		15.7	147	
			0.000			0.000			0.000
Students of color	73.2	4,267		13.4	780		13.4	780	
White	71.6	5,233		14.5	1,061		13.9	1,017	
Enrollment			0.000			0.000			0.000
Full time	65.8	4,232		17.0	1,096		17.2	1,104	
Part time	75.9	3,908		12.1	621		12.0	619	
Pell eligible			0.000			0.000			0.000
Yes	74.1	3,570		14.2	686		11.6	560	
No	73.2	5,159		13.9	979		13.0	914	
GPA range			0.000			0.000			0.000
3.0 or higher	45.4	1,639		35.7	1,287		18.9	681	
Below 3.0	84.2	4,404		7.4	389		8.4	438	
LASSI range			0.000			0.000			0.000
50-100	59.5	3,145		23.4	1,238		17.1	904	
0-49	74.2	2,744		9.5	350		16.3	602	
Status quo English placement			0.000			0.000			
1 level below college level	55.9	2,309		44.1	1,822				
2 levels below college level	97.2	4,441		2.8	128				
In the bump-up zone						0.000			
For both math and English				100.0	488				
For either math or English	35.1	953		58.6	1,591		6.3	171	
Sample size		10,894			2,082			2,026	

SOURCE: Demographic data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Distributions may not add to 100 percent because of rounding.

The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

To assess differences between the research groups, chi-square tests were used for categorical variables and two-tailed t-tests were used for continuous variables.

Enrollment status represents enrollment in the first semester. For one of the sites, this was determined based on credits attempted in the transcript data.

APPENDIX TABLE A.7 Enrolled into Gatekeeper English by Subgroup Among the Full Sample of Randomized Students

SUBGROUP (%)	SAMPLE	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL LOWER BOUND	90% CONFIDENCE INTERVAL UPPER BOUND	P-VALUE	P-VALUE DIFF. IN EFFECTS
Race/ethnicity								0.500
Asian	1,271	47.9	44.2	3.7	-0.7	8.2	0.169	
Black	2,423	44.2	39.1	5.0	1.9	8.1	0.008	
Hispanic	1,880	46.4	39.2	7.2	3.7	10.8	0.001	
White	8,089	48.6	45.5	3.1	1.3	4.8	0.003	
Other	1,079	49.7	45.8	3.9	-0.8	8.6	0.169	
								0.175
Students of color	6,653	46.4	41.2	5.2	3.3	7.0	0.000	
White	8,089	48.6	45.5	3.1	1.3	4.8	0.003	
Enrollment								0.901
Full time	7,202	61.6	57.4	4.1	2.3	5.9	0.000	
Part time	5,906	41.0	36.7	4.3	2.3	6.3	0.000	
Pell eligible								0.971
Yes	5,497	48.1	43.9	4.2	2.1	6.3	0.001	
No	7,733	47.2	43.0	4.2	2.5	6.0	0.000	
GPA range								0.037
3.0 or higher	4,166	60.6	53.2	7.4	5.0	9.9	0.000	
Below 3.0	5,868	44.4	41.0	3.4	1.3	5.4	0.006	
LASSI range								0.015
50-100	5,888	52.2	44.3	7.9	5.9	9.9	0.000	
0-49	4,129	46.0	42.8	3.3	0.8	5.7	0.026	
Status quo English placement								0.027
1 level below college level	3,579	40.3	27.7	12.6	10.1	15.1	0.000	
2 levels below college level	2,226	27.2	19.7	7.4	4.5	10.4	0.000	
In the bump-up zone								0.000
For both math and English	485	72.3	38.8	33.5	26.5	40.4	0.000	
For either math or English	2,926	60.1	45.7	14.4	11.5	17.4	0.000	
Sample size (total = 17,203)		10,476	6,727					

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.
Distributions may not add to 100 percent because categories are not mutually exclusive.
The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

APPENDIX TABLE A.8 Passed Gatekeeper English by Subgroup Among the Full Sample of Randomized Students

SUBGROUP (%)	SAMPLE	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL LOWER BOUND	90% CONFIDENCE INTERVAL UPPER BOUND	P-VALUE	P-VALUE DIFF. IN EFFECTS
Race/ethnicity								0.408
Asian	1,271	35.8	33.9	1.9	-2.5	6.3	0.473	
Black	2,423	27.5	25.2	2.3	-0.6	5.1	0.188	
Hispanic	1,880	30.0	25.1	5.0	1.6	8.3	0.014	
White	8,089	36.2	34.2	1.9	0.3	3.6	0.057	
Other	1,079	31.4	33.1	-1.8	-6.3	2.8	0.523	
								0.755
Students of color	6,653	30.4	28.1	2.4	0.6	4.2	0.028	
White	8,089	36.2	34.2	1.9	0.3	3.6	0.057	
Enrollment								0.180
Full time	7,202	43.4	42.4	1.1	-0.8	2.9	0.346	
Part time	5,906	29.0	25.8	3.2	1.4	5.0	0.004	
Pell eligible								0.462
Yes	5,497	30.4	28.8	1.6	-0.4	3.6	0.185	
No	7,733	36.1	33.4	2.8	1.1	4.4	0.007	
GPA range								0.023
3.0 or higher	4,166	50.6	45.7	4.9	2.4	7.4	0.001	
Below 3.0	5,868	26.0	25.5	0.6	-1.3	2.4	0.623	
LASSI range								0.392
50-100	5,888	37.4	34.0	3.5	1.5	5.4	0.003	
0-49	4,129	30.1	28.2	1.9	-0.3	4.2	0.163	
Status quo English placement								0.065
1 level below college level	3,579	27.2	20.1	7.1	4.8	9.4	0.000	
2 levels below college level	2,226	16.5	13.2	3.3	0.8	5.8	0.029	
In the bump-up zone								0.007
For both math and English	485	52.4	31.9	20.5	13.0	27.9	0.000	
For either math or English	2,926	44.4	37.0	7.4	4.5	10.3	0.000	
Sample size (total = 17,203)		10,476	6,727					

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.

Distributions may not add to 100 percent because categories are not mutually exclusive.

The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

APPENDIX TABLE A.9 Enrolled into Gatekeeper Math by Subgroup Among the Full Sample of Randomized Students

SUBGROUP (%)	SAMPLE	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL LOWER BOUND	90% CONFIDENCE INTERVAL UPPER BOUND	P-VALUE	P-VALUE DIFF. IN EFFECTS
Race/ethnicity								0.280
Asian	1,271	21.8	17.4	4.3	0.7	8.0	0.053	
Black	2,423	14.6	11.1	3.5	1.4	5.7	0.007	
Hispanic	1,880	18.0	13.4	4.6	1.9	7.3	0.005	
White	8,089	17.8	16.4	1.3	0.0	2.7	0.107	
Other	1,079	16.4	14.4	2.0	-1.6	5.6	0.358	
Students of color	6,653	17.2	13.5	3.7	2.2	5.1	0.000	0.049
White	8,089	17.8	16.4	1.3	0.0	2.7	0.107	
Enrollment								0.556
Full time	7,202	24.1	21.4	2.7	1.1	4.3	0.005	
Part time	5,906	13.7	11.7	2.0	0.6	3.4	0.022	
Pell eligible								0.093
Yes	5,497	15.6	12.1	3.5	2.0	5.0	0.000	
No	7,733	18.4	17.0	1.4	0.0	2.8	0.093	
GPA range								0.009
3.0 or higher	4,166	24.7	19.1	5.6	3.5	7.7	0.000	
Below 3.0	5,868	13.3	11.7	1.6	0.2	3.0	0.056	
LASSI range								0.001
50-100	5,888	18.9	12.6	6.3	4.8	7.8	0.000	
0-49	4,129	13.9	12.3	1.6	-0.1	3.2	0.127	
Status quo math placement								0.000
1 level below college level	4,126	25.4	17.6	7.8	5.7	9.8	0.000	
2 levels below college level	4,579	11.5	11.2	0.3	-1.2	1.8	0.736	
In the bump-up zone								0.000
For both math and English	485	41.4	12.2	29.3	22.4	36.1	0.000	
For either math or English	2,926	27.9	17.7	10.1	7.6	12.6	0.000	
Sample size (total = 17,203)		10,476	6,727					

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.
Distributions may not add to 100 percent because categories are not mutually exclusive.
The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

APPENDIX TABLE A.10 Passed Gatekeeper Math by Subgroup Among the Full Sample of Randomized Students

SUBGROUP (%)	SAMPLE	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL LOWER BOUND	90% CONFIDENCE INTERVAL UPPER BOUND	P-VALUE	P-VALUE DIFF. IN EFFECTS
Race/ethnicity								0.720
Asian	1,271	16.4	13.6	2.9	-0.5	6.2	0.156	
Black	2,423	9.2	7.8	1.4	-0.4	3.2	0.200	
Hispanic	1,880	9.3	7.9	1.5	-0.7	3.6	0.256	
White	8,089	12.3	12.0	0.3	-0.9	1.5	0.690	
Other	1,079	10.9	9.9	1.1	-2.0	4.2	0.572	
								0.222
Students of color	6,653	10.9	9.3	1.5	0.3	2.7	0.035	
White	8,089	12.3	12.0	0.3	-0.9	1.5	0.690	
Enrollment								0.853
Full time	7,202	16.1	15.2	0.9	-0.4	2.3	0.262	
Part time	5,906	9.2	8.4	0.7	-0.5	1.9	0.314	
Pell eligible								0.140
Yes	5,497	10.1	8.3	1.8	0.5	3.1	0.021	
No	7,733	12.3	12.1	0.2	-1.0	1.4	0.780	
GPA range								0.004
3.0 or higher	4,166	18.2	14.3	3.9	2.0	5.8	0.001	
Below 3.0	5,868	7.4	7.2	0.1	-1.0	1.3	0.837	
LASSI range								0.019
50-100	5,888	12.9	9.6	3.4	2.0	4.7	0.000	
0-49	4,129	9.2	8.7	0.5	-0.9	2.0	0.532	
Status quo math placement								0.000
1 level below college level	4,126	17.2	12.5	4.7	2.8	6.5	0.000	
2 levels below college level	4,579	7.0	7.6	-0.6	-1.8	0.7	0.453	
In the bump-up zone								0.001
For both math and English	485	26.2	8.2	18.0	11.9	24.1	0.000	
For either math or English	2,926	18.3	12.7	5.5	3.3	7.7	0.000	
Sample size (total = 17,203)		10,476	6,727					

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.
Distributions may not add to 100 percent because categories are not mutually exclusive.
The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

APPENDIX TABLE A.11 College-Level Credits Accumulated by Subgroup Among the Full Sample of Randomized Students

SUBGROUP (%)	SAMPLE	PROGRAM GROUP	CONTROL GROUP	DIFFERENCE	90% CONFIDENCE INTERVAL LOWER BOUND	90% CONFIDENCE INTERVAL UPPER BOUND	P-VALUE	P-VALUE DIFF. IN EFFECTS
Race/ethnicity								0.477
Asian	1,271	14.28	14.03	0.26	-0.87	1.39	0.706	
Black	2,423	10.47	10.04	0.43	-0.29	1.16	0.326	
Hispanic	1,880	11.50	10.60	0.91	0.01	1.80	0.095	
White	8,089	15.47	15.42	0.06	-0.42	0.54	0.845	
Other	1,079	12.61	13.26	-0.65	-1.87	0.58	0.384	
								0.447
Students of color	6,653	11.84	11.47	0.37	-0.10	0.84	0.198	
White	8,089	15.47	15.42	0.06	-0.42	0.54	0.845	
Enrollment								0.229
Full time	7,202	19.05	19.26	-0.21	-0.71	0.30	0.498	
Part time	5,906	10.71	10.45	0.27	-0.14	0.67	0.281	
Pell eligible								0.940
Yes	5,497	12.45	12.32	0.13	-0.38	0.64	0.667	
No	7,733	14.94	14.78	0.17	-0.33	0.66	0.580	
GPA range								0.800
3.0 or higher	4,166	18.80	18.48	0.31	-0.40	1.03	0.468	
Below 3.0	5,868	10.39	10.21	0.18	-0.32	0.68	0.550	
LASSI range								0.399
50-100	5,888	13.63	13.38	0.25	-0.28	0.78	0.437	
0-49	4,129	11.58	10.93	0.66	0.07	1.24	0.065	
Status quo math placement								0.068
1 level below college level	4,126	14.74	14.03	0.71	0.04	1.38	0.081	
2 levels below college level	4,579	11.79	12.08	-0.29	-0.89	0.31	0.431	
In the bump-up zone								0.003
For both math and English	485	16.61	12.11	4.50	2.43	6.57	0.000	
For either math or English	2,926	16.65	16.19	0.46	-0.36	1.29	0.355	
Sample size (total = 17,203)		10,476	6,727					

SOURCE: Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: Rounding may cause slight discrepancies in sums and differences.
Distributions may not add to 100 percent because categories are not mutually exclusive.
The p-value indicates the likelihood that the estimated impact (or larger) would have been generated by an intervention with zero true effect.

APPENDIX

B

How to Read the Area Under the Curve Plots for Receiver Operator Curves (AUC ROC)

To model the relationship between various assessment measures, we relied on the simplest statistical approach for a binary measure of success—a logistic regression. We also investigated whether we could improve predictive performance with two machine learning algorithms—LASSO (Least Absolute Shrinkage and Selection Operator) and Random Forest. However, neither of these machine learning approaches provided improvements in the predictive performance—perhaps because the total number of measures was not sufficiently large for them to add value, or because there were no meaningful interaction effects, or because the nature of the relationships between the predictors and outcomes is generally linear. Therefore, the plots in this appendix focus on logistic regression models. Findings from predictive models built with machine learning can be found in Appendix C.

We present a set of figures that plot the true positive rates and false positive rates across different thresholds in the predicted likelihoods of success in the college-level course.

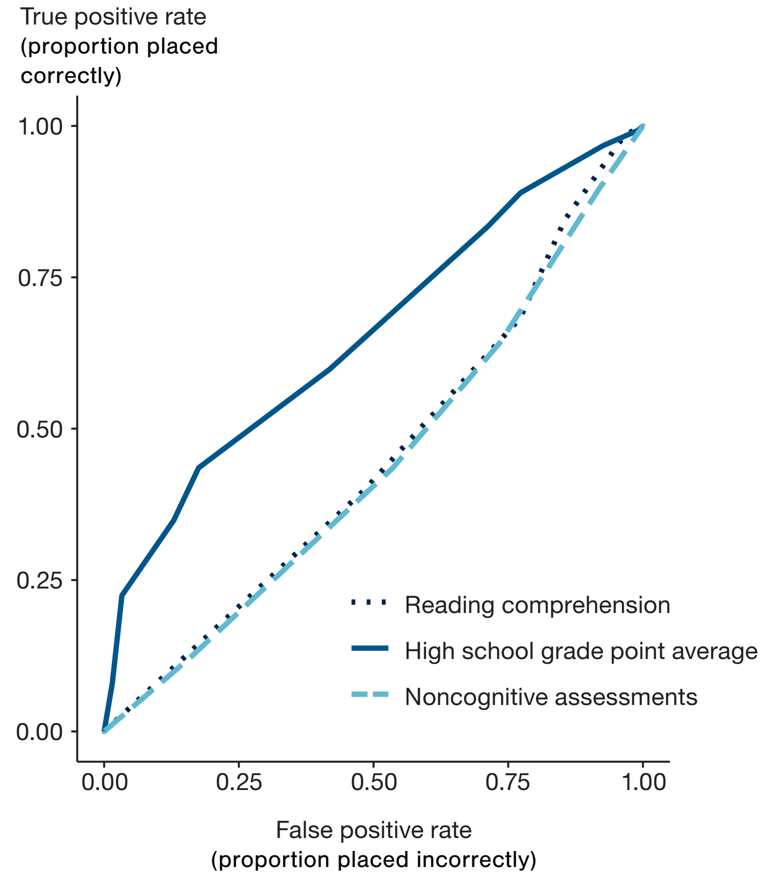
- The true positive rate: Among students who would succeed in a college-level course, this is the proportion correctly placed.
- The false positive rate: Among students who would not succeed in a college-level course, this is the proportion incorrectly placed.

Each line represents a combination of predictors in a logistic regression model. Each point on a line represents the true positive rate (Y-axis) and the false positive rate (X-axis) for a threshold. In this way, we can see all the trade-offs between the true positive rate and false positive rate. For example, a low threshold will recommend that a large portion of students be placed in the college-level course; however, many of these students will not succeed. Points for low thresholds are in the upper right corner of the plot. On the other hand, a high threshold will recommend that few students are placed in the college-level course; however, most of these students will succeed. Points for high thresholds are in the lower left corner of the plot. In the middle of these curves are points that are associated with more moderate thresholds.

Predictive models or decision-making procedures that are no better than a random coin flip will fall along a straight, 45-degree line from the lower left corner to the upper right corner. Curves that pull farther away from that 45-degree line and reach closer to the upper left corner are generally associated with predictive models or decision-making procedures that are doing a better job at correctly placing students. Such curves will have higher true positive rates and lower false positive rates. An empirical summary of the predictive performance is the area under the curve (AUC). These are often referred to receiver operator curves (ROCs), so this empirical summary is the AUC ROC. All AUC ROC values are summarized in Appendix C.

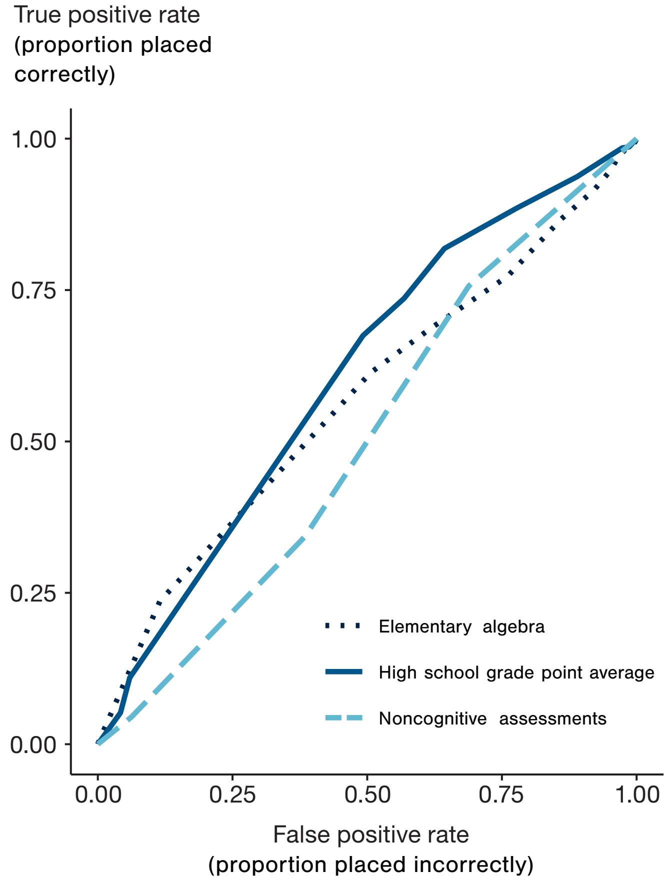
The plots in this appendix exclude predictor sets that combined multiple measures—such as ACCUPLACER and grade point average—because the performance of these models was not much different from the models with a single predictor. However, the AUC ROC values for these models can be found in Appendix C.

FIGURE B.1 Predictive Performance of Reading Comprehension Scores



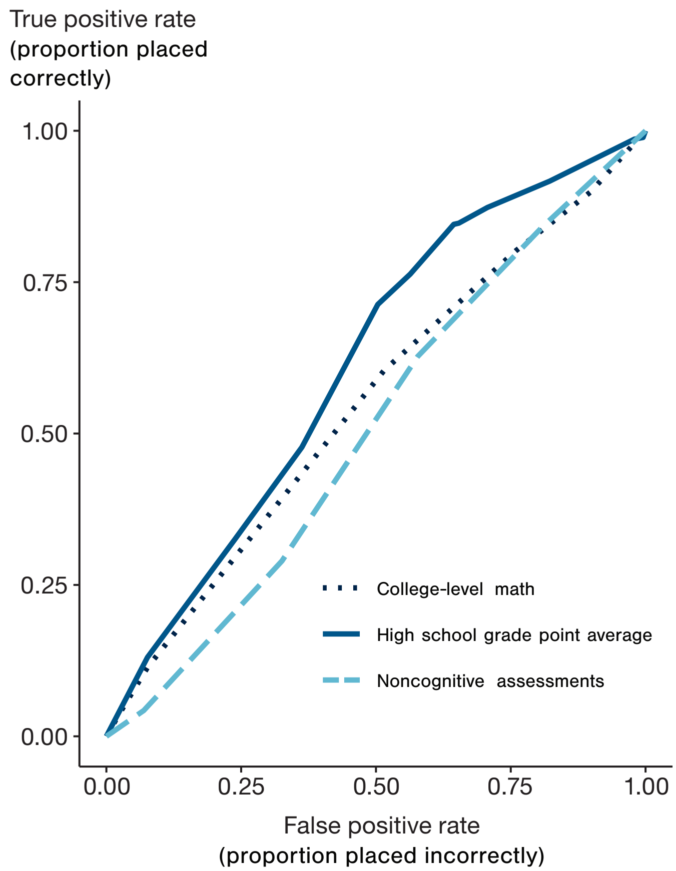
SOURCE: Placement and Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

FIGURE B.2 Predictive Performance of Elementary Algebra Scores



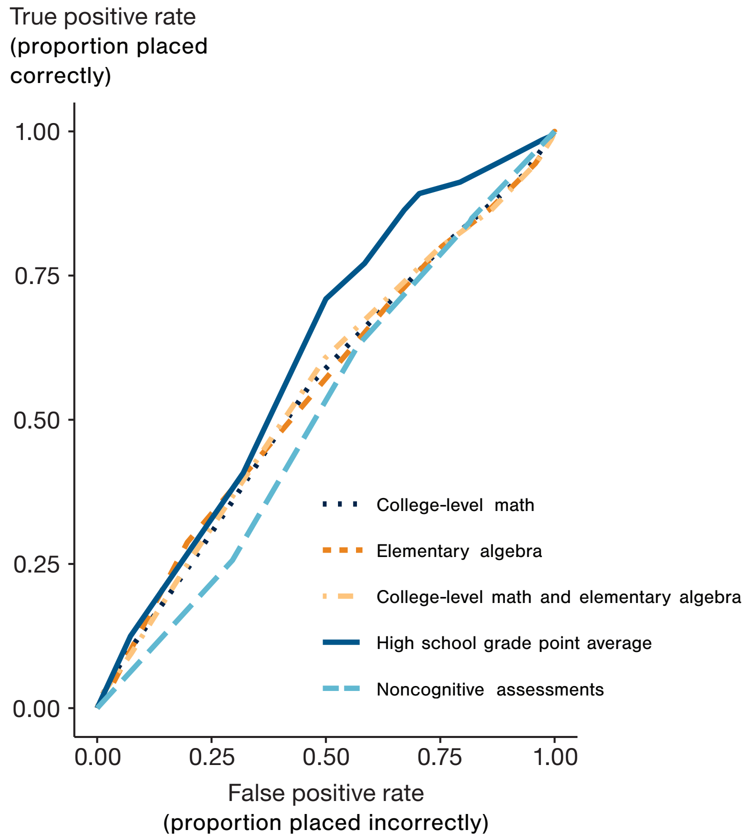
SOURCE: Placement and Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

FIGURE B.3 Predictive Performance of College-Level Math Scores



SOURCE: Placement and Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

FIGURE B.4 Predictive Performance of Elementary Algebra and College-Level Math Scores



SOURCE: Placement and Transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

APPENDIX

C

Predictive Analysis Tables

How to Interpret the Tables

To model the relationship between various assessment measures and passing a college-level course, a statistical approach for a binary measure of success was used—a logistic regression (a general linear model). The research team also attempted to improve predictive performance with two machine learning algorithms—LASSO (Least Absolute Shrinkage and Selection Operator) and Random Forest. A sample partitioning approach (cross validation) was used to avoid producing a model that overfits the data. Each model was fit on data from all but one college, and then used to obtain predictions for the students in the left-out college. This was repeated for all colleges until every student had a predicted likelihood.

The tables in this appendix summarize the predictive performance of all models using the area under the curve (AUC) of a receiver operator curve. Each plot shows the true positive rate on the Y-axis, or the proportion of correctly placed students among those who would succeed in a college-level course, and the false positive rate on the X-axis, or the proportion of incorrectly placed students among those who would not succeed in a college-level course. The area under the curve of each plot is the AUC. AUC values closer to 1 indicate better predictive performance, and values of 0.5 are no better at predicting an outcome than a coin flip.

APPENDIX TABLE C.1 Area Under the Curve (AUC) for the Reading Comprehension Sample

MODEL	PREDICTOR SET	AUC ROC
GLM	College-level math	0.547
	College-level math & high school GPA	0.620
	College-level math & high school GPA & noncognitive assessments	0.611
	College-level math & noncognitive assessments	0.552
	High school GPA	0.601
	Noncognitive assessments	0.511
LASSO	College-level math & high school GPA	0.620
	College-level math & high school GPA & noncognitive assessments	0.611
	College-level math & noncognitive assessments	0.551
Random Forest	College-level math & high school GPA	0.623
	College-level math & high school GPA & noncognitive assessments	0.623
	College-level math & noncognitive assessments	0.557

SOURCE: Placement and transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: ROC = receiver operator curve, GPA = grade point average.

TABLE C.2 Area Under the Curve (AUC) for the Elementary Algebra Sample

MODEL	PREDICTOR SET	AUC ROC
GLM	Elementary algebra	0.570
	Elementary algebra & high school GPA	0.604
	Elementary algebra & high school GPA & noncognitive assessments	0.610
	Elementary algebra & noncognitive assessments	0.569
	High school GPA	0.620
	Noncognitive assessments	0.501
LASSO	Elementary algebra & high school GPA	0.603
	Elementary algebra & high school GPA & noncognitive assessments	0.609
	Elementary algebra & noncognitive assessments	0.569
Random Forest	Elementary algebra & high school GPA	0.608
	Elementary algebra & high school GPA & noncognitive assessments	0.621
	Elementary algebra & noncognitive assessments	0.572

SOURCE: Placement and transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: ROC = receiver operator curve, GPA = grade point average.

TABLE C.3 Area Under the Curve (AUC) for the College-Level Math Sample

MODEL	PREDICTOR SET	AUC ROC
GLM	College-level math	0.547
	College-level math & high school GPA	0.620
	College-level math & high school GPA & noncognitive assessments	0.611
	College-level math & noncognitive assessments	0.552
	High school GPA	0.601
	Noncognitive assessments	0.511
LASSO	College-level math & high school GPA	0.620
	College-level math & high school GPA & noncognitive assessments	0.611
	College-level math & noncognitive assessments	0.551
Random Forest	College-level math & high school GPA	0.623
	College-level math & high school GPA & noncognitive assessments	0.623
	College-level math & noncognitive assessments	0.557

SOURCE: Placement and transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: ROC = receiver operator curve, GPA = grade point average.

TABLE C.4 Area Under the Curve (AUC) for the Elementary Algebra and College-Level Math Sample

MODEL	PREDICTOR SET	AUC ROC
GLM	College-level math	0.548
	College-level math & elementary algebra	0.563
	College-level math & elementary algebra & high school GPA	0.623
	College-level math & elementary algebra & high school GPA & noncognitive assessments	0.615
	College-level math & elementary algebra & noncognitive assessments	0.560
	Elementary algebra	0.566
	High school GPA	0.620
	Noncognitive assessments	0.507
LASSO	College-level math & elementary algebra & high school GPA	0.623
	College-level math & elementary algebra & high school GPA & noncognitive assessments	0.615
	College-level math & elementary algebra & noncognitive assessments	0.560
Random Forest	College-level math & elementary algebra & high school GPA	0.629
	College-level math & elementary algebra & high school GPA & noncognitive assessments	0.622
	College-level math & elementary algebra & noncognitive assessments	0.584

SOURCE: Placement and transcript data provided by Anoka-Ramsey Community, Century, Minneapolis Community and Technical, Normandale, and Madison colleges.

NOTES: ROC = receiver operator curve, GPA = grade point average.

REFERENCES

Anzelone, Caitlin, Michael Weiss, and Camielle Headlam, with Xavier Alemañy. 2020. *How to Encourage College Summer Enrollment: Final Lessons from the EASE Project*. New York: MDRC.

Barnett, Elisabeth A., Peter Bergman, Elizabeth Kopko, Vikash Reddy, Clive Belfield, and Susha Roy. 2018. *Multiple Measures Placement Using Data Analytics: An Implementation and Early Impacts Report*. New York: Community College Research Center, Teachers College, Columbia University.

Barnett, Elisabeth A., Elizabeth Kopko, Dan Cullinan, and Clive R. Belfield. 2020. *Who Should Take College-Level Courses?: Impact Findings from an Evaluation of a Multiple Measures Assessment Strategy*. New York: Community College Research Center, Teachers College, Columbia University; and MDRC.

Belfield, Clive, and Peter M. Crosta. 2012. "Predicting Success in College: The Importance of Placement Tests and High School Transcripts." CCRC Working Paper No. 42. New York: Community College Research Center, Teachers College, Columbia University.

Bickerstaff, Susan, Elizabeth Kopko, Erika B. Lewy, Julia Raufman, and Elizabeth Zachry Rutschow. 2021. *Implementing and Scaling Multiple Measures Assessment in the Context of COVID-19*. New York: MDRC.

Carson, Andrew D. 2012. "Predicting Student Success from the LASSI for Learning Online (LLO)." *Journal of Educational Computing Research* 45, 4: 399–414.

Chen, Xianglei, Lesa R. Caves, Joshua Pretlow, Samuel Austin Caperton, Michael Bryan, and Darryl Cooney. 2020. *Courses Taken, Credits Earned, and Time to Degree: A First Look at the Postsecondary Transcripts of 2011–12 Beginning Postsecondary Students*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Chen, Xianglei, Michael A. Duprey, Nichole Smith Ritchie, Lesa R. Caves, Daniel J. Pratt, David H. Wilson, Frederick S. Brown, and Katherine Leu. 2020. *High School Longitudinal Study of 2009 (HSLS:09): A First Look at the Postsecondary Transcripts and Student Financial Aid Records of Fall 2009 Ninth-Graders*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Cullinan, Dan, and Elisabeth A. Barnett. 2021. *Ascendium Multiple Measures Assessment Project*. New York: MDRC and CCRC.

REFERENCES *(continued)*

Cullinan, Dan, Elisabeth A. Barnett, Elizabeth Kopko, Andrea Lopez, and Tiffany Morton. 2019. *Expanding Access to College-Level Courses: Early Findings from an Experimental Study of Multiple Measures Assessment and Placement*. New York: MDRC and CCRC.

Cullinan, Dan, Elisabeth A. Barnett, Alyssa Ratledge, Rashida Welbeck, Clive Belfield, and Andrea Lopez. 2018. *Toward Better College Course Placement: A Guide to Launching a Multiple Measures Assessment System*. New York: MDRC.

Dadgar, Mina, Linda Collins, and Kathleen Schaefer. 2015. *Placed for Success: How California Community Colleges Can Improve Accuracy of Placement in English and Math Courses, Reduce Remediation Rates, and Improve Student Success.* San Rafael, CA: The RP Group Career Ladders Project.

Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology* 92, 6: 1087.

Rugsaken, Kris T., Jacqueline A. Robertson, and James A. Jones. 1998. "Using the Learning and Study Strategies Inventory Scores as Additional Predictors of Student Academic Performance." *NACADA* 18, 1: 20–26.

Rutschow, Elizabeth Zachry, Maria S. Cormier, Dominique Dukes, and Diana E. Cruz Zamora. 2019. *The Changing Landscape of Developmental Education Practices: Findings from a National Survey and Interviews with Postsecondary Institutions*. New York: Community College Research Center, Teachers College, Columbia University; and Center for the Analysis of Postsecondary Readiness, MDRC.

Scott-Clayton, Judith. 2012. "Do High-Stakes Placement Exams Predict College Success?" Working Paper No. 41. New York: Community College Research Center, Teachers College, Columbia University.

Weinstein, Claire Ellen, David R. Palmer, and Taylor W. Acee. 2016. *LASSI User's Manual: Learning and Study Strategies Inventory*. Dunedin, FL: H&H Publishing.

ABOUT MDRC

MDRC, A NONPROFIT, NONPARTISAN SOCIAL AND EDUCATION POLICY RESEARCH ORGANIZATION, IS COMMITTED TO finding solutions to some of the most difficult problems facing the nation. We aim to reduce poverty and bolster economic mobility; improve early child development, public education, and pathways from high school to college completion and careers; and reduce inequities in the criminal justice system. Our partners include public agencies and school systems, nonprofit and community-based organizations, private philanthropies, and others who are creating opportunity for individuals, families, and communities.

Founded in 1974, MDRC builds and applies evidence about changes in policy and practice that can improve the well-being of people who are economically disadvantaged. In service of this goal, we work alongside our programmatic partners and the people they serve to identify and design more effective and equitable approaches. We work with them to strengthen the impact of those approaches. And we work with them to evaluate policies or practices using the highest research standards. Our staff members have an unusual combination of research and organizational experience, with expertise in the latest qualitative and quantitative research methods, data science, behavioral science, culturally responsive practices, and collaborative design and program improvement processes. To disseminate what we learn, we actively engage with policymakers, practitioners, public and private funders, and others to apply the best evidence available to the decisions they are making.

MDRC works in almost every state and all the nation's largest cities, with offices in New York City; Oakland, California; Washington, DC; and Los Angeles.