Multiple Measures Assessment: Improving Course Placement in Two-Year Institutions

A THESIS
SUBMITTED TO THE FACULTY
OF THE UNIVERSITY OF MINNESOTA
BY

Corwin Talbert

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Professor David R. Johnson

May, 2017

Acknowledgements

First, I would like to thank God, my creator and redeemer, who has sustained and strengthened me by his Grace. His mercies are new each morning. Second, I would like to thank my family who has offered nothing but love, encouragement, and support during this time. I must especially thank my wife, Caitlyn, who has been a source of strength for me and whose own academic pursuits have been an inspiration. I am forever grateful for our life together and for her continued support. Special thanks to my adviser, Dr. David Johnson, for his guidance in the development of this thesis, and Dr. Ernest Davenport, for his helpful feedback on drafts. Lastly, I would like to thank my colleagues at Minneapolis Community and Technical College for their work and collaboration on this project, of which this thesis is a product.

Abstract

Each year, many students arrive on community college campuses needing to take developmental, pre-college coursework prior to beginning their programs of study. Often, the only factor used in assessing students' college readiness is a placement exam that students take when they first apply. There is increasing evidence that these exams misplace substantial proportions of students, delaying and possibly deterring them from degree attainment. It has been suggested that incorporating variables related to students' prior academic achievement and noncognitive traits can aid in placement. This thesis explored the current state of assessment for course placement in English and mathematics and the need to develop more holistic systems to place students properly. These issues were examined in the context of Minneapolis Community and Technical College, a diverse, two-year institution in Minneapolis, Minnesota. The current placement system at the college was analyzed, which relies heavily on the ACCUPLACER, a placement testing suite. Two of the four tests were found to be poor predictors of course success and exhibited large score disparities among racial and ethnic groups. The predictive validity of high school GPA, high school rank, and the Grit Scale were analyzed. High school GPA proved to be a consistently strong predictor and improved diagnostic accuracy by reducing severe misplacement. Lastly, based on these findings, recommendations were made for the implementation of a multiple measures placement system.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Course Placement and Developmental Education in Community Colleges

Community colleges in the United States are traditionally nonselective, open-access institutions: anyone with a high school diploma or GED can enroll. However, students must be deemed "college ready" before being allowed to take credit-bearing, transferrable coursework, and to determine their readiness, colleges administer a placement exam when they are first admitted. These placement exams measure students' proficiency in English, reading, and mathematics. The use of these instruments is widespread: according to a representative survey, all two-year public institutions reported using some type of mathematics test for assessment purposes and 94% reported using a reading test (Fields & Parsad, 2012).

Often, these placement exams are the sole assessment used to determine a students' readiness for college-level coursework. The reasons institutions use placement testing revolve around feasibility. Scott-Clayton, Crosta, and Belfield (2014) state that "the affordability and efficiency of the screening tool itself are clearly important, particularly for large institutions that may need to process thousands of entrants within a matter of weeks" (p. 373). For many students, the result of assessment is placement into remedial or developmental courses.[1] These courses do not bear college credit and are considered pre-college skill development courses. They are designed to raise students' proficiencies to college standards for success in mainstream courses.

The proportion of students who are assessed as needing developmental coursework in at least one area is considerable. Attewell, Lavin, Domina, and Levey (2006) found that more than half of community college students take at least one developmental course. Although traditionally thought of as just a minor setback in students' academic career, there is increasing evidence that placing a semester or more behind in English and mathematics can have drastic effects on

[1] The literature uses both "remedial" and "developmental" to describe English, reading, and mathematics courses below the college level. I will use "developmental" throughout this paper.

students' chances of graduating or transferring (Attewell et al., 2006; Bailey, 2008; Bailey, Jeong, & Cho, 2010).

Placement Test Validity and the Promise of Multiple Measures

In recent years, there has been increased attention on the validity of placement exams to accurately differentiate between students needing developmental coursework and those who are college ready. There is strong evidence that these tests are susceptible to error, resulting in significant proportions of students being "underplaced" and "overplaced." Underplacement is defined as assigning students to developmental courses who could have succeeded at the college level; overplacement is assigning students to college-level courses who are predicted to fail there. One can think of these as false negative and false positive results respectively. Scott-Clayton et al. (2014) found that severe misplacements are common, with underplacement much more common than overplacement. Underplacement has the negative impact of delaying students' academic pursuits and eventual degree attainment by requiring them to take additional coursework they do not need. Overplacement may overmatch students with the course curriculum, leading them to fail or withdraw.

It has been proposed that measures such as high school transcript information have additional and sometimes stronger predictive power than traditional placement tests and that implementing a multiple measures approach to assessment may increase the overall accuracy of course placement. Even the placement test publishers themselves recommend using more than just test scores to make placement decisions (College Board, 2016). High school GPA in particular has the potential of measuring not only prior educational achievement but also other characteristics such as effort and motivation (cited in Noble & Sawyer, 2004). There is also a wealth of literature on noncognitive assessments that predict college outcomes. Instruments that attempt to measure such constructs as grit, perseverance, motivation, and learning behavior have yet to be formally examined for placement, but there is increasing policy and legislative interest to do so (Barnett, 2017; Burdman, 2012; Ngo & Kwon, 2014). Utilizing data about students' prior academic achievement and noncognitive traits could increase the accuracy of course placement, reducing the rates of misplacement and increasing course success and eventual college completion.

Background on Minneapolis Community and Technical College

The current study is applied in nature. It seeks to examine current assessment practices and offer recommendations for placement system improvements at Minneapolis Community and Technical College (MCTC). MCTC is a diverse, urban two-year institution located in Minneapolis, Minnesota that serves approximately 12,000 students annually. Over half of enrollees are students of color and approximately 46% are low income. It is a member of Minnesota State, Minnesota's public state college system, which is the fourth largest system in the U.S., with 30 colleges, 7 universities, and 54 campuses statewide. (Minneapolis Community and Technical College, 2017).

Purpose Statement

The purpose of this study is to evaluate the current state of assessment for course placement at MCTC, to explore additional information not currently in use that could prove valuable for use in a multiple measures system, and to propose a modified assessment system that incorporates these additional measures to improve assessment policy and outcomes at the college. A further purpose of this study is to inform an ongoing project. I am an employee of the college, and this past fall, our institution along with five other Minnesota State community colleges elected to participate in a study sponsored by the education and social policy research organization MDRC and the Community College Research Center (CCRC) to develop and implement a multiple measures assessment system. The two-year project entails undertaking an initial historical data analysis; vetting, developing, and executing a pilot of the new system; and eventually conducting a larger, randomized control trial to evaluate its effectiveness. The findings from this study will directly steer project proposals and decision-making.

Lastly, being an employee of the college, I feel a strong sense of advocacy for the students. I have worked at MCTC for over four years in different roles and have met many of our students on their educational journeys. I have witnessed the consequences of placement testing first-hand and desire for this research to drive decision-making for the benefit, success, and equity of all MCTC students, especially students of color who may be most negatively impacted by traditional placement practices.

Positionality

My academic and professional background inform and influence my views on the nature of social science research and, consequently, the current study's methodology and analytical framework. In short, I would describe myself as a cautious post-positivist. I reject wholesale the tenets of positivism and believe that all human knowledge is ultimately contingent on the knower and is always fallible, imperfect, and in need of constant revision. In practice, I am a pragmatist; I believe the choice of method should be informed by the ultimate aims of the research project. The research presented in this study reflects that approach. Simply put, it is a quantitative study because these methods are most useful for analyzing the predictive validity of assessment tools and for setting new policy for decision rules and cutoffs. Lastly, it is important to acknowledge a certain level of comfort and preference for quantitative research methods, as this has encompassed much of my time both as a graduate student and research analyst.

Research Questions

This study was guided by the following research questions:

1. What is the current state of assessment for course placement at Minneapolis Community and Technical College?

2. What additional student information beyond placement test scores is useful for increasing placement accuracy?
3. How can this additional information be incorporated into a multiple measures assessment system?

Methodology

This study analyzed the predictive validity of placement tests, high school transcript information, and a noncognitive questionnaire for use in a multiple measures assessment system. First, goodness-of-fit statistics were computed for these variables, using tests of correlation to measure the predictive power and statistical relationship of different variables with course success in English and mathematics. Second, as prediction is ultimately the goal of any assessment instrument used for placement, measures of diagnostic accuracy were conducted. This entailed calculating assignment error rates, specifically what is deemed the “severe error rate” (SER) developed by Scott-Clayton et al. (2014). Lastly, further analysis and discussion of severe error rates under different assignment rules and cutoffs was undertaken to help determine how best to incorporate additional measures into placement decisions.

Chapter 2. Literature Review

The Purpose of Developmental Education and the Concept of College Readiness

To gain a proper understanding of the state of placement testing in community colleges, it is helpful to give an overview of concepts of college readiness and developmental education. The principal purpose of tools such as placement exams is to differentiate between who is college ready and who needs developmental coursework, so an introduction to these concepts will help frame the subsequent discussion and analysis. Often, students arriving on two-year college campuses possess skills judged to be too weak in the areas of English, reading, and mathematics for success in the mainstream courses institutions offer. Institutions address this need with developmental education. These courses are designed to strengthen basic skills: for instance, the ability to write clear, organized prose; the ability to read a text critically; and the ability to reason quantitatively. College-level courses assume the student possesses these skills and can put them to adequate use to overcome the rigors of the curricula.

What does it mean to be college ready? There are various definitions and indicators as one might expect, but they can be broadly categorized as those focusing on predictors, like prior high school achievement, and those focusing on outcomes, like the ability to avoid developmental coursework upon arrival to college (Maruyama, 2012). These two notions are essentially the same regarding placement: the prior academic characteristics and achievement a student possesses at the point of enrollment should give an accurate picture as to his or her propensity and readiness for college. Operationally, college readiness is the level of preparation necessary to enroll and succeed in a credit-bearing course at a postsecondary institution (Conley, 2007). According to Conley (2012), this notion should encompass four key areas: cognitive strategies, content knowledge, learning skills and techniques, and transition knowledge and skills. Any assessment tool or set of tools should have the ability to measure one or more of these constructs to properly estimate students' likelihood for success at the college level.

The Landscape of Developmental Education

What are the overall proportions of students who are referred to developmental courses, and what are their educational outcomes? There are a number of reports from both state and national governmental agencies that provide estimates of developmental education enrollment and success (Fergus, 2016; Parsad, Lewis, & Greene, 2003). There are also formal research studies that explore developmental course-taking using longitudinal datasets. These studies provide estimates of not only enrollment and demographics but also the simulated effects of remediation on graduation and transfer (Attewell et al., 2006; Bailey, 2008; Bailey et al., 2010).

One of the most oft-cited reports on broad estimates of developmental education enrollment comes from the National Center for Education Statistics (NCES) (Parsad et al., 2003). This report looked at enrollment across the country in the fall of 2000 and found that 42% of freshman at public two-year institutions enrolled in at least one developmental English, reading, or mathematics course. Of those enrolled in developmental courses at two-year institutions, 37% spent less than one year taking these courses, 53% spent one year, and 10% spent more than one year. Other more recent studies show the proportion of students taking developmental courses has increased.

Given the current study's context, it is worthwhile to get an estimate of developmental education enrollment in Minnesota. The Minnesota Office of Higher Education, pursuant to Minnesota law, has been tasked with reporting on the state of developmental education enrollment of recent high school graduates (Fergus, 2016). Their annual report, entitled *Getting Prepared*, provides summaries of students who enrolled in developmental courses at colleges across the state within two years of graduating from a Minnesota high school. The 2016 report found that 26% of 2013 public high school graduates enrolled in at least one developmental course within two years of graduation. Among those enrolled in developmental education courses, 85% were enrolled in Minnesota public two-year colleges of which MCTC is one.

Nearly half of 2013 high school graduates enrolling in two-year institutions took developmental courses. Regardless of institution type, students of color enrolled in developmental education courses at higher rates than White students, Black or African American students being the highest at 53%. Clearly, substantial proportions of recent high school graduates are being deemed unprepared for college in Minnesota. This evidence may speak to the disconnect between high school and college curricula; it may also speak to the assessment practices at colleges in Minnesota and the need to evaluate these to ensure students are placed properly.

There is a wealth of literature that attempts not only to estimate enrollment trends but also the effects and outcomes of students assigned to developmental coursework. Attewell et al. (2006) analyzed the National Educational Longitudinal Study, known as the NELS:88, a project of the National Center on Educational Statistics (NCES). This was a nationally representative sample of eighth grade students who later provided postsecondary enrollment information. Rich college transcript data allowed the researchers to analyze students' course-taking behavior. Of these students, 40% took at least one developmental course, with mathematics being the most common subject taken.

Developmental course enrollment was much more common for NELS students attending two-year institutions than four-year colleges and universities: 58% of two-year college students took at least one developmental course compared to 31% of students at nonselective four-year colleges. Of these community college students, 44% took between one and three developmental courses, and 14% took more than three courses. Of developmental students in two-year institutions, only 28% graduated within 8.5 years compared to 43% of students who did not take developmental coursework. However, there is competing evidence that students who enrolled and subsequently passed developmental coursework had better educational outcomes than those who did not take any developmental coursework. The most pertinent finding for the current study is that many students with *poor* high school records did not take developmental coursework, while substantial numbers of students with *strong* high school performance did take these courses. If high school achievement is related to college performance, this finding may speak to the low predictive power of the assessment tool or tools for course placement used at these colleges and the need for more accurate measures for assessing college readiness.

Another study that examined a large, longitudinal dataset was conducted by Bailey et al. (2010), which looked at developmental education enrollment and outcomes from data collected for the Achieving the Dream: Community Colleges Count initiative, a national initiative created to improve persistence and completion for community college students. Their study broadens the discussion of developmental referral and enrollment, focusing attention on the entire *sequence* of developmental education courses. Often, especially in mathematics, there is more than one course below the college level; students may place two or three courses below college level and need to complete the entire sequence of developmental courses before being eligible to take mainstream courses. The dataset in the study contained over 250,000 first-time students who began college between fall 2003 and fall 2004 and were followed for three full academic years. An important variable in the dataset for the purposes of their study was an indicator of developmental education referral and the *level* at which they were referred (i.e., where in the developmental sequence a student was placed). Overall, 59% of students were referred to developmental math—24% one level below the college level, 16% two levels below, and 19% three or more levels; 33% were referred to developmental reading. Maybe the most staggering of these numbers is that roughly one in five students were referred to a developmental math sequence that was three or more terms long; assuming the college was on a semester schedule, students placed here would have needed to take and complete a year and a half or more of developmental mathematics before being allowed to enroll in the college-level, credit-bearing course.

The study also highlighted developmental sequence completion rates. Overall, 33% of students assigned to any level of developmental math completed their sequence, and 46% of those referred to any developmental reading completed. For students three or more levels below the college level in math, only 17% completed their sequence. Lastly, only 20% of those assigned to developmental math completed a college-level math course. What stands out from this study is the sheer complexity of student referral, enrollment, and completion. One would tend to think that developmental education sequences work within a strong policy structure, but the authors discovered that many students enrolled in courses contrary to their placement. They found that barely a majority of students actually followed their referrals, often enrolling in the college-level course even though they placed lower. And those who did enroll in the developmental sequences to which they were referred took a wide variety of pathways to get to the college-level courses. This evidence may hint that students' perceptions of their own readiness is at odds with their assessment results, and that more can be done to ensure accurate assessment for course placement not only to determine college readiness but also the optimal *level* within developmental sequences.

Assessment for Course Placement at Two-Year Institutions

Just as there are national reports and in-depth research studies on the state of developmental education, there exist analogous literature for assessment and placement testing within higher education (Belfield & Crosta, 2012; Burdman, 2012; Fields & Parsad, 2012; Scott-Clayton, 2012; Scott-Clayton et al., 2014). These reports and studies attempt to give a picture of the landscape of placement testing and its effects on student outcomes with the hopes of informing practice. The summary below attempts to synthesize the most recent and significant findings in this expanding area of educational policy research.

The National Assessment Governing Board oversees and sets policy for the National Assessment for Educational Progress (NAEP), and in the fall of 2011 they were tasked with reporting on the state of placement testing nationally (Fields & Parsad, 2012). Their survey questions asked about the types of tests and assessments used for placement and where college-level thresholds were set for these tests. As was reported above, 100% of two-year public institutions reported using some type of standardized mathematics test for course placement and 94% used a standardized reading test.

One especially relevant finding from Fields and Parsad (2012) for the current study is that they asked respondents about their use of any other criterion above and beyond standardized tests for placement. Overall, 21% of all institutions and 27% of two-year public institutions used other criteria beyond mathematics tests for course placement. For reading it was even less: only 13% of institutions stated using criteria beyond standardized tests for placement and only 19% of two-year institutions. Evidently, the use of multiple measures for course placement is infrequent. Further, only 8% of two-year public institutions reported using high school grades (including GPA) for course placement in mathematics and only 4% in reading. If high school transcript information proves a valuable predictor for the purposes of course placement, which as will be shown below, it does, there exists a major opportunity for research to test and validate it for use. If less than one in ten two-year institutions incorporate high school data for placement decisions, and these data could greatly improve the quality of assessment, it is imperative that institutions start adopting practices that incorporate these measures.

The Predictive Validity of Placement Exams and the Need for Multiple Measures

As analyzing the validity and accuracy of placement tests is central to the current study, it is important to investigate into influential research in this area. Given the nearly widespread use of these placement exams at community colleges across the country, how accurately do they measure students' college readiness in reading, writing, and mathematics? That is, do placement exams do an adequate job in predicting future performance (i.e., do they have predictive validity)? And is there other information about students beyond placement exams that can aid in the assessment process?

Predictive Validity and Background on the ACCUPLACER. Sawyer (1996) states that traditionally, validating tests for use in educational assessment emphasizes statistical relationships between test scores and relevant criteria; the key statistic for measuring this has been the correlation coefficient. Validity for placement tests is typically measured by the correlation of test scores and relevant course grades. Another estimate of a test's predictive validity is a measure of accuracy rates, i.e., how good the instrument predicts those who will be successful and unsuccessful in the college-level course. The accuracy rate is the frequency of correct classifications estimated by individuals' probability of success (Mattern & Packman, 2009). It is noteworthy that much of the predictive validity studies done to calculate these statistics for placement exams are produced by the test publishers themselves (Hughes & Scott-Clayton, 2011).

The ACCUPLACER is a suite of up to six computer-adaptive tests in reading comprehension, writing and sentence skills, and mathematics from basic arithmetic to advanced algebra. The tests are anywhere from 13 to 20 questions in length and are untimed, each taking approximately 30 minutes to complete (College Board, 2004). As MCTC uses the ACCUPLACER, the review of predictive validity literature will focus primarily on this exam suite. A comprehensive predictive validity study of the ACCUPLACER was performed by Mattern & Packman (2009), who used a meta-analytical approach to estimate correlation coefficients and accuracy rates. They synthesized 47 validity studies from 2006 to 2011 with 17 unique institutions, 14 of them from two-year colleges. Sample-size weighted correlations were calculated for the six ACCUPLACER exams (Arithmetic, Elementary Algebra, College-Level Mathematics, Reading Comprehension, Sentence Skills, and WritePlacer) with success defined as earning a B or higher or C or higher in the relevant course. The observed biserial correlations ranged from .16 to .36 for B or higher, and the .10 to .32 for C or higher. Correlations across the board were higher for the B or higher criterion than the C or higher criterion and tended to be stronger for the mathematics tests than the English or reading tests. The percent correctly placed (i.e., the accuracy rate) ranged from 73% to 84%. If one takes the strongest single correlation from any of the ACCUPLACER tests, and squares the statistic to get a broad estimate of the amount of variance in course success that is explained by the exam, the best these tests do is explain roughly 13% of the variance in course success. As will be discussed below, in applied settings with local college systems, a number of these tests tend to do even worse than this, making the case for exploring additional measures for course placement even stronger.

The Value of High School Transcript Information. There is an abundance of evidence that information about students' academic achievement and course-taking in high school are strong predictors of future success at the postsecondary level (Geiser & Santelices, 2007; Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008; Long, Iatarola, & Conger, 2009; Ngo & Kwon, 2014; Noble & Sawyer, 2004; Sawyer, 2013). High school GPA in particular is a rich statistic in predicting future academic success. It is accumulated over four years as opposed to a standardized test score that is determined in a few hours. In a study by Geiser and Santelices (2007) of almost 80,000 students admitted to the University of California, high school GPA was consistently the best predictor of both freshman grades and performance, as well as four-year college outcomes. They also found that high school GPA had less of an adverse impact on students of color and underrepresented populations than standardized tests—a finding that is corroborated in the present study. Unfortunately, there has been little applied use of high school transcript information for course placement in community colleges. But, there is burgeoning policy research advocating for its use, which holds promise for systematic change.

Analysis of Placement Accuracy from the Community College Research Center. Over the past seven years, the Community College Research Center (CCRC), housed at the Teacher's College of Columbia University, has published a number of influential working papers and journal articles calling into question the validity of placement exams (Belfield & Crosta, 2012; Hughes & Scott-Clayton, 2011; Scott-Clayton, 2012; Scott-Clayton et al., 2014). The latest publication (Scott-Clayton et al., 2014) provides the most comprehensive research to date by the CCRC in this area, as it made use of both of the two large community college system datasets presented in Scott-Clayton (2012) and Belfield & Crosta (2012). The 2014 study examined the placement accuracy of ACCUPLACER and COMPASS test scores and then estimated the incremental value of high school transcript information in making placement decisions. Rather than focusing on correlation coefficients or other goodness-of-fit statistics, their analysis involved calculating rates of diagnostic accuracy. They estimated how often the placement exam assessed students as developmental when they could have succeeded at the college level (underplacement), and how often the exam placed them at the college level when they were predicted to fail there (overplacement). Rather than computing accuracy rates, which may vary depending on how "success" is defined (e.g., is it B or higher, C or higher, or even D or higher, as a D still earns a student credit for most courses?), they decided to narrow their analysis on error rates and what they call the "severe error rate." Many potential cutoffs can generate similar accuracy rates while the underplacement and overplacement rates can vary drastically (Sawyer, 1996). This gives additional information to practitioners who may prefer to minimize one type of error over the other.

The severe error rate (SER) combines the proportion of those predicted to earn an A or B in the college-level course but instead placed into developmental with the proportion of students placed into college level but were predicted to fail or withdraw there. They also calculated the predicted success rate of those placed into the college-level course under a given assignment rule, as well as the overall proportions assigned to developmental and college level (what they call the "remediation rate").[2] They found that in both English and mathematics, using high school GPA alone as a placement tool resulted in fewer proportions of students being severely misplaced, with error rate reductions ranging from 12% to 30%. Both underplacement and overplacement rates were reduced, and at the same time the estimated success rates in these college-level courses increased. Stated in another way, when they held the remediation rate constant (i.e., assigned the same proportions of students to developmental and college-level courses as the traditional placement test cutoffs did) but instead used high school GPA for course placement, they reduced both the rates of severe under- and overplacement as well as increased the predicted course success rates in the college-level course. For some of the courses, the SER was further reduced by using high school GPA in conjunction with test scores. Given these promising findings for how high school transcript information may ameliorate the negative effects of placement tests, the current study takes a similar approach to evaluate MCTC's placement system.

Noncognitive Factors. An additional measure which may prove useful for course placement is the employment of noncognitive assessments. Although the term *noncognitive* in educational psychology is nebulous and imprecise (Duckworth & Yeager, 2015), here I take it to mean certain qualities of a student distinct from cognitive abilities that have potential to impact academic success. Much of the research on noncognitive factors focuses on college performance. These studies are interested in the determinants and predictors of success in postsecondary education above and beyond traditional measures of standardized tests and prior high school performance. They explore a broad range of psychological factors and constructs related to psychosocial and study skill behavior in order to ultimately uncover and describe, as one study puts it, "the third pillar of academic success" (Crede & Kuncel, 2008, p. 425).

[2] The current study utilizes these same metrics under different assignment rules on local data, so the exact methodology for calculating these will be given later below.

One major area of research on noncognitive traits involves study and learning behavior. Crede and Kuncel (2008) conducted a large meta-analysis of different study habit, skill, and attitude inventories to investigate the predictive validity of such measures on academic performance in college. They analyzed data from a vast number of inventories that attempt to reveal these constructs, but the majority of the data came from two assessments in particular: the Survey of Study Habits and Attitudes (SSHA) and the Learning and Study Skills Inventory (LASSI). They found that these inventories were generally independent of both high school performance and admissions exams. Further, *study motivation* and *study skills* constructs were consistently the strongest predictors of college GPA, and *academic anxiety* was found to be negatively related to performance. They concluded that study skills, habits, and attitudes more than any other set of noncognitive traits studied to date improve the prediction of college performance.

Another large meta-analysis of psychosocial and noncognitive factors was conducted by Robbins et al. (2004), which produced subsequent studies (Allen, Robbins, Casillas, & Oh, 2008; Komarraju, Ramsey, & Rinella, 2013; Peterson, Casillas, & Robbins, 2006; Porchea, Allen, Robbins, & Phelps, 2016) and the construction of the Student Readiness Inventory (Le, Casillas, Robbins, & Langley, 2005). The original meta-analysis synthesized literature related to a wide range of constructs including social influence and engagement, academic motivation and goals, and notions of self-worth. They collected and synthesized a total of 476 correlations (197 with a retention criterion and 279 with a GPA criterion) from 109 studies. Most of the factors analyzed in the study were found to correlate positively with both college retention and GPA. After controlling for the traditional predictors of high school achievement and standardized test scores, *academic goals*, *self-efficacy*, and *academic-related skills* were the strongest predictors of retention, and *achievement motivation* was found to be the strongest predictor of GPA. They concluded by proposing three higher order constructs from the literature: *motivation*, *study and learning skills*, and *social engagement*.

Lastly, the literature on grit and perseverance has recently received considerable attention in education and may prove to be another facet of college readiness worth examining for course placement. The notion of grit in educational psychology was first systematically studied by Duckworth, Peterson, Matthews, and Kelly (2007). They defined grit as resilient effort towards long-term goals despite setbacks and adversity. Dissatisfied with the scales available at the time to measure this construct, they developed their own self-report questionnaire called the Grit Scale. They found that it accounted for variance in educational outcomes such as college GPA and educational attainment and demonstrated incremental predictive validity beyond traditional cognitive measures. In another study, Strayhorn (2014) found that grit was positively related to college grades for Black students attending a predominantly White institution and that it added predictive validity above and beyond high school GPA and ACT scores. Wolters and Hussain (2015) found that one of the two dimensions of grit, *perseverance of effort*, was a significant predictor of college students' current academic performance and that grit was related to students' self-regulated learning. Ultimately, the relationship between grit and college academic performance is unknown as the literature is quite limited. The current study seeks to shed further light on this construct in academic settings.

The placement test publishers themselves are aware of this growing research field and have begun to incorporate noncognitive components into their testing suites. For instance, the ACCUPLACER has developed what they call "multiple weighted measures," where answers to a number of locally-determined background questions give students a weighted boost in their final scores (College Board, 2013).[3] They believe this can help ameliorate the difficult burden placed

[3] In fact, MCTC has used this feature in the ACCUPLACER to boost students' test scores for the past few fiscal years. As the current study's purpose was not to evaluate the effectiveness of multiple weighted measures, only raw test scores were used in the analysis below.

on placement exams to correctly classify students near the course placement cutoffs. This "decision zone" as they call it, is where answers to noncognitive questions related to motivation or time management could boost scores of students on the cusp of placement into the higher course.

Ngo and Kwon (2014) provide one of the few analyses on the type of multiple measures boosts described above. Analyzing math placement and enrollment data from the Los Angeles Community College District, they looked at the effect of boosts students received from answering background questions related to prior high school performance, college plans, and motivation. Overall, only 4% of students were placed into the higher-level course due to the multiple measures boost they received. They found that the use of multiple measures marginally increased the number of underrepresented students being placed into a higher level of math and that boosted students performed no worse than their non-boosted peers. They also found that both high school GPA and responses about previous math performance were positive predictors of course success; unfortunately, they were unable to parse out the singular effect of the noncognitive questions related to college plans and motivation. They concluded that the validity of these measures has yet to be explored in the context of placement decisions that entail test score adjustments and that further research is needed.

The Prevalence of Multiple Measures Assessment. Given the ample evidence of the strength of high school transcript information and certain noncognitive factors in predicting college performance, it is discouraging that the vast majority of two-year institutions do not use this information to make placement decisions. As was stated above, only around one in four institutions use any criteria beyond placement exams to assist in course placement. Burdman (2012) provides an overview of how different state college systems are beginning to rethink their traditional assessment practices.

The California community college system has used multiple measures in some form since 1991. Forty-five of the state's 112 colleges embed questions in their computerized assessment system about previous academic experience in certain subjects and self-reported high school statistics (Venezia, Bracco, & Nodine, 2010). It appears that because of the large amount of autonomy as to how individual campuses implement these multiple measures, there is little systematic research on its impact (see Ngo & Kwon, 2014 for one example). However, recent efforts to evaluate assessment at these colleges shows promise. The RP Group, a non-profit organization committed to increasing student success in community colleges in California, is currently leading a large project to implement a statewide multiple measures placement tool (RP Group, 2017). The project's objectives are to create a data warehouse for multiple measures and pilot new placement systems based on high school transcript information and other variables.

For many other state systems, the use of multiple measures is nascent at best. Few states have operational K-16 data systems for collecting relevant high school data in real-time. As of 2012, North Carolina had the technical capacity and was in the midst of reviewing the research and making recommendations for policy, but no concerted effort had been underway. New Jersey decided to use high school information to help make decisions for students near the cutoff scores for placement into college-level courses. Connecticut has legislation that requires the use of multiple measures at its state colleges and universities, and pilots are being conducted (Burdman, 2012). It is clear that multiple measures assessment is gaining traction in numerous states across the country. Given the opportunities it offers to increase placement accuracy, it is incumbent upon institutions and college systems to explore their assessment data to inform practice.

Gaps in the Literature

There are several gaps in the literature that this study addresses. First, little to no research to date has analyzed placement accuracy for different *levels* of developmental courses. The focus has been on the all-important developmental/college level divide, without attention placed on placement decisions within developmental course sequences. As significant proportions of MCTC students test into developmental math especially, it is critical that this study analyze the assessment accuracy within these course levels in order to make the largest impact.

Second, previous research has not provided substantive detail as to *how* policymakers and practitioners implement a multiple measures assessment system. The literature shows the value of high school transcript information and noncognitive assessments' ability to predict college performance above and beyond test scores but lacks the practical details to take advantage of these measures for placement. How do college staff incorporate both high school transcript data and placement test scores to place students? Where should thresholds for these metrics be set to ensure accuracy? I explore these questions in the discussion chapter below and offer recommendations for academic professionals.

Lastly, from the review of noncognitive literature, there is no formal research that examines using noncognitive assessments for placement decisions. Virtually all the studies that focus on psychosocial and study skills constructs use a generic criterion of college performance such as retention or GPA for validation rather than individual course performance. By analyzing the Grit Scale's relationship with English and math course success, this study adds to the limited research in this area. The analysis below, which is applied and evaluative in nature, addresses these gaps in an effort to add to the knowledgebase for course placement and assessment at two-year institutions.

Chapter 3. Methods

Introduction

This study set out to answer three primary research questions. First, what is the current state of assessment for course placement at Minneapolis Community and Technical College? Second, what additional student information beyond placement test scores is useful for increasing placement accuracy? Lastly, how can this additional information be incorporated into a multiple measures assessment system? To answer the first two questions, I evaluated the predictive validity of ACCUPLACER test scores, high school GPA and rank, and Grit Scale scores by correlating these with success in English and mathematics courses. This gave a broad picture of the power of different predictors. I then looked at the ACCUPLACER's ability to accurately place students through calculating severe error rates and whether these errors were reduced by incorporating high school GPA as an assessment tool. Lastly, I discussed ways in which assessment policymakers and practitioners can set cutoffs and decision rules through further analysis of these error rates.

Analytical Framework

As is clear from the literature review, much of the research on assessment for course placement relies heavily on quantitative methods. There are a number of articles, however, that take qualitative or mixed methods approaches. These studies explore such topics as students' experiences and attitudes of the placement process (Venezia et al., 2010) and policy analysis as it differs across systems (Melguizo, Kosiewicz, Prather, & Bos, 2014). But given the current study's scope and ultimate aims, quantitative methods are the most adequate tools for the task. Additional measures being considered for placement need to be conducive to quantitative analysis, as they must work in a system that handles thousands of student applicants every term in an automated process. Thus, it is reasonable to apply quantitative statistical methods to estimate their usefulness and suitability.

Validity Metrics

Correlation coefficients. The predictive validity of placement tests and other measures is typically estimated by computing correlation coefficients between the predictor (e.g., test scores, high school GPA, etc.) and course grades, which are dichotomized as success (defined as earning a C or higher or B or higher) and non-success. In two studies by the College Board, the publisher of the ACCUPLACER and SAT exams, they calculated estimates of validity in this way (Kobrin et al., 2008; Mattern & Packman, 2009). The correlation coefficient is a measure of the linear relationship of two variables (Fox, 2016). It can be thought of as a standardized covariance (as it ranges from −1.0 to 1.0) that shows how negatively or positively two variables are related. Typically, an absolute value of .1 is considered a small correlation; .3 is medium; and .5 or higher is large (Cohen, 1992).

This study first looked at the correlation between a number of predictors and course grades in English and mathematics. I correlated the four ACCUPLACER exams administered at the college and their respective course or courses: the Reading Comprehension test and English Composition, and the three mathematics tests (Arithmetic, Elementary Algebra, and College-Level Math) and three mathematics courses—two developmental and one college level. I then looked at several other predictors: high school GPA, high school rank, and the mean score on the 8-item Short Grit Scale (Duckworth & Quinn, 2009). The Grit Scale was administered to students immediately following the ACCUPLACER during the 2014-2015 academic year, and the mean of the questionnaire was computed per the authors' recommendations. I correlated these variables with the same courses outlined above to compare their relative strength to the ACCUPLACER. For all the correlations computed, I dichotomized course grades into 1 = C or higher, 0 = below a C. Correlations where one of the variables is dichotomous are called point-biserial correlations; if the variable is dichotomized artificially and has an underlying continuity, which was the case here with course grades, one can compute a biserial correlation (Field, Miles, & Field, 2012). Both point-biserial and biserial correlations are presented below.

Range Restriction. An important additional statistical adjustment is called for with placement test scores. As the test publishers make explicit (Kobrin et al., 2008; Mattern & Packman, 2009), there is a restriction of range with these scores, because the institutions already have cutoff scores in place. For instance, at MCTC, only students with a score of 78 or higher on the Reading Comprehension test can register for college-level English. Now in practice, there is a small proportion of students who gain access to courses for which they do not have a passing score. Students can obtain a waiver or exemption into the course if they provide certain evidence; there is also an ACT cutoff score that can get them in the course. Nonetheless, for the vast majority of students, only a score above the cutoff gets them in the course. That means for English Composition, the typical range of scores of students in the course is not 20 to 120 but rather 78 to 120. Figure 1 illustrates this feature of the data. This range restriction underestimates the correlation coefficients because the linear relationship between the two variables is obscured by the reduced variability in the predictor. The subset of scores above the cutoff only gives a partial picture of how the two variables are related.

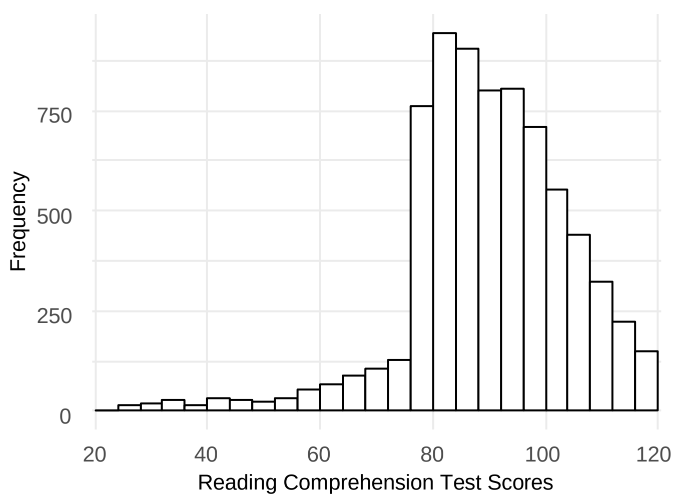

FIGURE 1. Histogram of Test Scores of English Composition Students

There are a number of ways to correct for range restriction, but maybe the most traditional and commonly used is called *Thorndike case 2* (Wiberg & Sundström, 2009). The formula uses the correlation of the restricted sample, and the standard deviation of the predictor variable (i.e., the test score) in the restricted sample and in the unrestricted sample to provide an estimate of the correlation for the population. The formula is as follows:

$$r_{XY} = \frac{S_X r_{xy}}{\sqrt{S_X^2 r_{xy}^2 + s_x^2 - s_x^2 r_{xy}^2}},$$

where

r_{xy} is the observed correlation between X (test score) and Y (course success) in the restricted sample.

s_x is the estimated standard deviation of X in the *restricted* sample.

S_x is the estimated standard deviation of X in the *unrestricted* sample.

r_{XY} is the estimated corrected correlation between X and Y.

To estimate the SD of the unrestricted sample, I used the SD of all test takers in my dataset. This correction gave a slight boost to the correlations compared to the raw *r* computed. I did this only for the ACCUPLACER scores, as these were the only measures where there was a cutoff policy in place.

Estimating Severe Error Rates. The other common procedure for estimating a measure's predictive validity for course placement is through the notion of diagnostic accuracy—that is, what percentage of students were correctly and incorrectly placed. As mentioned above, the current study estimated severe error rates rather than accuracy rates per the reasoning and recommendations of Scott-Clayton et al. (2014). The SER combines the proportion of students predicted to earn an A or B in the college-level course but were placed into developmental and the proportion of students predicted to earn an F or W (withdraw) in the college-level course but were placed there. By focusing on error rates, one can determine what alternative assessment measures make the *least severe misclassifications*, which is more practical for policymakers who often need to satisfy a diverse set of stakeholders. One can also see the differing proportions of under- and overplacement errors, which aids in decision-making given stakeholders' preferences. The first step in estimating the SER is to develop as rich a predictive model as possible from the data for both the A or B outcome variable and the F or W outcome variable. Scott-Clayton et al. (2014) used probit regression for these models, but I employed logistic regression as this is a more common method for predicting a dichotomous outcome variable, especially in the field of placement test validity (see Morgan & Michaelides, 2005; Noble & Sawyer, 2004; Sawyer, 1996, 2007). Scott-Clayton et al. (2014) advised creating an estimation sample and a prediction sample. The estimation sample is a restriction of only those students who took the course in question. For each course, I ran the following logistic regression models:

$$\ln\left(\frac{p(\text{AorB} = 1)}{1 - p(\text{AorB} = 1)}\right) = \beta_0 + \beta_1\text{Test} + \beta_2\text{HS} + \boldsymbol{\beta_3}\mathbf{Dem} + \varepsilon,$$

$$\ln\left(\frac{p(\text{ForW} = 1)}{1 - p(\text{ForW} = 1)}\right) = \beta_0 + \beta_1\text{Test} + \beta_2\text{HS} + \boldsymbol{\beta_3}\mathbf{Dem} + \varepsilon,$$

where

Test was the relevant ACCUPLACER test score.

HS was HS GPA.

Dem was a vector of demographic variables; this may have included age, gender, and/or low-income status.

These models differed depending on which course was under consideration and what showed as statistically significant and aided in the prediction of the outcome variable. I ran these two logistic regressions for four courses separately: English Composition (the college-level English course, called ENGL 1110), College Algebra (the college-level math course, called MATH 1110), Intermediate Algebra (one level below college-level math, called MATH 0080), and Introductory Algebra (two levels below college-level math, called MATH 0070).

Next, I computed the predicted probabilities from both regression equations for *all* students with the relevant data who may or may not have placed into the course in question. This group is called the prediction sample. The probability of being severely *underplaced* is equal to the predicted probability from the first equation for students who placed below the level in question; otherwise it is 0 (because they were necessarily not severely underplaced). The probability of being severely *overplaced* is equal to the predicted probability from the second equation for students who placed at or above the level in question; otherwise it is 0. For the English sample, that level was the college level, but it differed for the math estimations. The estimate of the severe error rate for an individual is the sum of these two predicted probabilities. The SER for the sample as a whole then is simply the mean of these individual predicted probabilities.

The SER was first estimated using only the ACCUPLACER for placement (the college's current practice). I then simulated using alternative tools to see how the SER changed. To approximate this, I needed to hold the overall assignment rates fixed. For instance, if under the ACCUPLACER, 52% of students placed into college-level English, then when I switched to using HS GPA, I still needed to assign 52% of students to college-level English. To do this, I just found the percentile cutoff that placed the same proportions of students to the different levels as before. For multiple predictors (like using test scores and HS GPA together), I used the ordered predicted probabilities of a multiple logistic regression. Lastly, I estimated the predicted course success rate by taking the mean of predicted probabilities for students above the theoretical cutoff from a logistic regression model with only the assessment tool(s) in question as the predictor(s).

Institutional Context and Data

Sample and Data Elements. I analyzed first-time, degree-seeking students who enrolled at MCTC from 2007 to 2016 (N = 20,368). There were some students who enrolled at the college under a different admission status prior to their degree-seeking status; these students were excluded from the final sample. At the college, students can retest for any of the ACCUPLACER tests, so the highest score of each test prior to the start of their first term was used. When computing correlations and the SER estimates, I used only students who took the course as their first course in the relevant subject; if, for example, they took developmental English prior to taking college-level English, they are excluded from the college-level English estimates. I then pulled their demographic data, high school GPA and rank, and Grit Scale score if available.

ACCUPLACER Cutoff Scores. It is important to stress again that there are ACCUPLACER cutoff scores currently in place at the college. Table 1 shows where these current cutoffs are for placement; these have varied slightly in developmental math over time, and the placement estimates take this into account. The cutoff for placement into college-level English is a 78 or higher on the Reading Comprehension test, and the cutoff for placement into college-level math is a 50 or higher on the College-Level Math test. These college-level cutoffs are set at the Minnesota State system level, and individual institutions set cutoffs below these. Students have a few options for developmental coursework in English: students with scores between 60 and 77 can take Accelerated Fundamentals of Composition (ENGA 0900/1110), which allows them to complete their college-level English requirement in one semester, or students can take the traditional developmental English course, Fundamentals of English (ENGL 0900). For math, Intermediate Algebra (MATH 0080) has a cutoff of 76 on the Elementary Algebra test, and Introductory Algebra (MATH 0070) has a cutoff of 46 on the Elementary Algebra test or a 50 on the Arithmetic test. These are, respectively, one and two levels below the college level in math. Lastly, scores in the lowest ranges for both English and math place students in adult basic education (ABE) courses, which are taught through Minneapolis Public Schools

TABLE 1. Current ACCUPLACER Cutoff Scores and Placements

ACCUPLACER Test	Minimum Score	Maximum Score	Course	Placement
Reading Comprehension	78	120	ENGL 1110	College Ready
Reading Comprehension	60	77	ENGA 0900/1110	Developmental
Reading Comprehension	38	77	ENGL 0900	Developmental
Reading Comprehension	20	37	ABE	ABE
College-Level Math	50	85	MATH 1110	College Ready
Elementary Algebra	76	120	MATH 0080	Developmental
Elementary Algebra	46	75	MATH 0070	Developmental
Arithmetic	50	120	MATH 0070	Developmental
Arithmetic	20	49	ABE	ABE

ACCUPLACER Branching Profiles. The ACCUPLACER at MCTC also has branching profiles for administering the different tests. When students come to take the ACCUPLACER, they are given a set of background questions to determine which tests to administer. If they answer that they are a non-native English speaker, they are branched into the ESL ACCUPLACER tests; depending on how well they perform there, they may be administered the traditional Reading Comprehension test and math tests. With regard to math, the question "What is the highest-level math class you have completed?" branches them into starting with either the Arithmetic test or the Elementary Algebra test.[4] Depending on their score of their first administered math test, they may be given additional tests. They are branched into different tests depending on the cutoff scores delineated above. For instance, if they start in Arithmetic and score higher than a 50 (the cutoff into MATH 0070), they are then administered the Elementary Algebra test to see if they can place higher. The only way to be administered the College-Level Math test is if the student scores higher than a 76 on the Elementary Algebra test (i.e., the cutoff score for MATH 0080). Consequently, as will be apparent in some of the correlations below, only a small proportion of students even had College-Level Math test scores on file because they did

[4] If they answer "None" or "Basic Math (arithmetic)," they are administered the Arithmetic test. If they answer "Algebra I/Beginning Algebra," "Algebra II/Intermediate Algebra," "One year beyond Algebra II (Trig/College Algebra/Pre-Calculus)," or "Two years beyond Algebra II (Calculus/Analysis or higher)," they are administered the Elementary Algebra test.

not perform well enough on the other tests to be branched there. Overall, 89% of the entire sample had a Reading Comprehension test score on file; 82% had an Arithmetic score; 65% had an Elementary Algebra score; and only 13% had a College-Level Math score.

Demographics. Table 2 shows the demographics of the entire sample and the subsamples I draw from. It is clear that very few students had high school transcript information on file. I only display proportions with high school GPA, as this was the most frequent high school statistic available. Overall, only 14% of students had a high school GPA on record, and of those students, they tended to be younger: the average age of the entire sample was 24, but for those who had reading test scores and a high school GPA, it was 19, nearly 5 years younger. What is not shown here but is especially promising for the stakeholders at the college, is that in recent years, there has been an increase in collecting high school information. This bodes well if the institution decides to use these data for assessment purposes.

Another important highlight from the descriptive statistics is that 91% of students were not college ready in mathematics compared to just 44% for English. Only using the ACCUPLACER and the institution's cutoff scores, it was even higher: 96% of students tested developmental in mathematics and 49% in English. This leads to what is called the problem of extrapolation (Sawyer, 1996; Scott-Clayton et al., 2014). Because setting cutoff scores and computing accuracy rates and correlation coefficients uses only the subset of students who took the college-level course, there is a concern that one cannot sensibly extrapolate to those below the cutoff. This is, in part, the rationale for analyzing students who placed and took courses below the college level in math, namely, those in Introductory Algebra (two levels below) and Intermediate Algebra (one level below).

TABLE 2. Overall Demographics

		English Sample		Math Sample	
	All first-time, degree-seeking students	Reading Test Takers	Reading Test Takers with HS GPA	Math Test Takers	Math Test Takers with HS GPA
Sample Size	*20,368*	*18,131*	*2,556*	*19,233*	*2,644*
% Female	53.4	53.4	45.5	53.7	54.4
Race/Ethnicity					
% Student of Color	66.2	65.0	62.8	66.8	64.7
% Asian	4.9	4.3	8.4	4.8	8.8
% Black or African American	40.0	39.3	31.5	40.6	33.5
% Hispanic or Latino	10.2	9.7	11.7	10.2	11.6
% White	31.7	33.5	36.9	31.2	34.8
Avg. Age	23.9	23.8	18.9	23.9	19.0
% Low Income	69.5	70.2	64.5	70.6	66.1
% First-Generation	32.5	31.2	28.9	32.8	30.6
ACCUPLACER					
Avg. Reading Comprehension Score	75.9	75.9	76.4	75.8	76.1
Avg. Arithmetic Score	51.5	51.6	55.7	51.6	54.9
Avg. Elementary Algebra Score	51.5	50.3	55.0	51.5	54.9
Avg. College-Level Math Score	41.7	40.4	39.3	41.8	39.5
% Assigned to Developmental					
In English	43.8	49.4	42.3	45.1	44.1
In Math	91.0	93.6	91.6	96.3	95.8
In Either Subject	92.4	95.6	95.5	98.7	98.9
% with HS GPA	14.0	14.1		13.7	
Avg. HS GPA	2.6	2.5		2.6	
% with Grit	6.5	6.7	8.2	6.7	8.6
Avg. Grit	3.7	3.7	3.6	3.7	3.7

Underrepresented Students. Lastly, it is important to stress the level of diversity at the college. Nearly two-thirds of the students in the sample were students of color, with Black or African-American students comprising the largest single race/ethnicity group. Also, 70% of students overall were low income, as defined by eligibility for the Pell grant. Nearly one-third of students were first-generation college students.[5] There were slightly less students of color in the samples that had high school GPA: Asians and Hispanic students made up a larger proportion

[5] This the Minnesota definition, which is defined as a student whose parents never enrolled at a postsecondary institution

compared to the sample as a whole, and there was a smaller proportion of Black or African-American students and a greater proportion of White students.

One of the goals of this study was to look at the disparate impact of assessment on underrepresented groups like students of color. There were large disparities in placement results from the ACCUPLACER. For instance, 71% of Black or African-American students tested into developmental English as opposed to only 21% of White students. The disparity was much smaller for math, with 98% of Black or African-American students testing into developmental compared to 96% for White students, but this is because nearly all students tested below college level in math anyway. The real disparity in math was found in those who tested into adult basic education: 55% of Black or African American students tested into ABE in math compared to only 20% for White students.

I also computed the effect sizes (ES) for mean differences of the ACCUPLACER tests and high school GPA and rank by race/ethnicity. For calculating the ES, I used Cohen's *d*, which is simply the mean score for a given group subtracted by the mean score of the base group divided by the standard deviation (Cortina & Nouri, 2000). I chose to use the pooled standard deviation from these four groups (approximated by the square root of the mean squared error of a one-way ANOVA). Figure 2 shows these effect sizes with White students as the base group. One can see that both the Reading Comprehension test and the Arithmetic test had much larger effect sizes by race/ethnicity than the other two ACCUPLACER tests and the high school statistics. Cohen (1992) proposed the following operational definitions for the values of *d*: 0.2 is small, 0.5 is medium, and 0.8 is large.

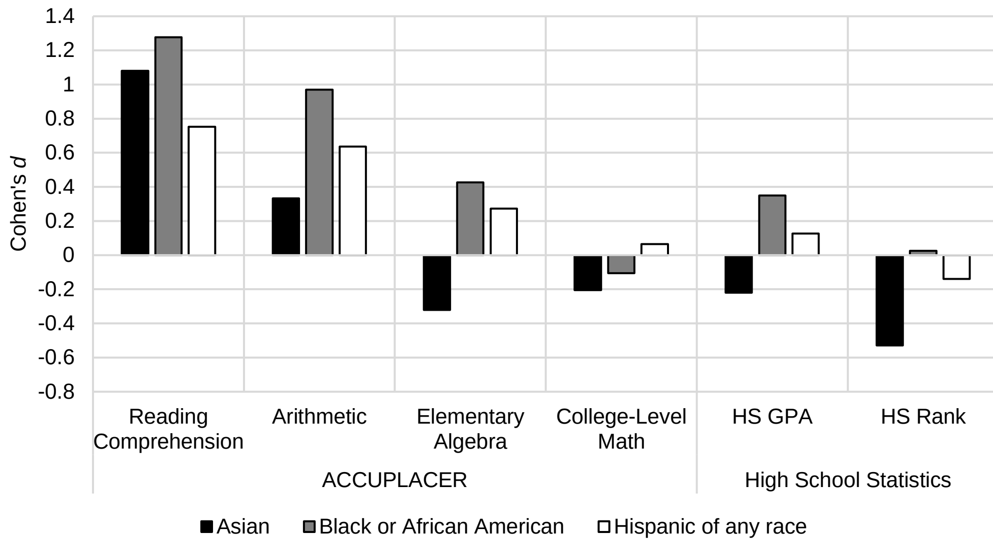

FIGURE 2. Effect Sizes of ACCUPLACER Tests and HS Statistics by Race/Ethnicity

For the Reading Comprehension test, these values were drastic: for both Asian and Black or African American students, *d* was extremely large, and for Hispanic students, it was medium-large. This indicates that on average, minority students performed much worse than White students on this test. These values remained medium-large to large for the Arithmetic test as well, especially for Black or African Americans. For the Elementary Algebra test and College-Level Math test, *d* was much smaller, and Asians outperformed Whites (negative *d* values). Both tests may have been prone to selection bias, however. Since only those who performed well on the lower ACCUPLACER tests were branched into the higher tests, both were a select subset of higher-performing students. This may be an indication that there were less racial and ethnic disparities in higher-performing test-takers.

High school statistics showed much smaller values of *d* as well. There existed no medium or large effect sizes except for a negative ES for Asian students and HS rank. However, there may have been some selection bias here too: from the descriptive statistics above, students with HS GPA had slightly higher average ACCUPLACER scores on three of the tests. They also had a slightly different demographic makeup. Nevertheless, HS GPA and HS rank appear to be less prone to racial and ethnic disparities and could help temper what appears to be an "assessment gap" at the college. The most commonly administered ACCUPLACER tests (Reading Comprehension and Arithmetic) had the largest race/ethnicity gaps, which gives an additional impetus to explore multiple measures to improve equity at the college.

Chapter 4. Results

Correlation Results

ACCUPLACER. Table 3 shows the results from the correlation tests. All of the ACCUPLACER tests used for placement into the related courses in English and Math were statistically significant. The raw biserial correlation coefficients for the Reading Comprehension and Arithmetic test, however, were quite small (.05 and .10 respectively); even after applying the correction for range restriction, they were still small and thus poor predictors of success in the courses for which they placed students. Conversely, the Elementary Algebra and CLM test had much larger correlations. The Elementary Algebra test was the best predictor as is evidenced by its correlations with both MATH 0070 and MATH 0080. The CLM test had a corrected biserial correlation of .19 but was only significant at the .01 level. The sample size for the CLM and MATH 1110 correlation was only 2% of the overall sample, which showed just how rare it was for students at MCTC to have both tested college ready in math and subsequently enrolled in College Algebra (MATH 1110).

High School Statistics. For high school statistics, one can see that both HS GPA and HS rank were much better predictors across all courses except for HS rank and MATH 1110. HS GPA was five times larger for ENGL 1110 than the Reading Comprehension test (.35 versus .07) and three times larger for MATH 0070 than the Arithmetic test (.39 versus .12). In fact, HS GPA was a better single predictor than ACCUPLACER in every comparison except the Elementary Algebra test for MATH 0070. When comparing HS GPA and HS rank, the correlation for HS GPA was larger for every course and was much more available in the data, which is evident from the sample sizes. Further, HS GPA and HS rank were highly correlated with each other (Pearson's r = .83), and when used together in multiple regression models, HS GPA tended to be both more predictive and statistically significant. Thus, subsequent analysis focused exclusively on HS GPA.

TABLE 3. Correlation Results for ACCUPLACER, HS GPA, HS Rank, and Grit Scale

ACCUPLACER

Test	*Course*	*n*	r_{pb}	$r_{b(correction)}$	p_b	SD_b
RC	ENGL 1110	7,259	.04	.05 (.07)	<.001	.01
CLM	MATH 1110	365	.14	.19 (.27)	.006	.07
EA	MATH 0080	1,175	.14	.18 (.31)	<.001	.04
EA	MATH 0070	2,032	.20	.25 (.40)	<.001	.03
AR	MATH 0070	1,821	.08	.10 (.12)	<.001	.03

HS GPA

Course	*n*	r_{pb}	r_b	p_b	SD_b
ENGL 1110	1,304	.27	.35	<.001	.04
MATH 1110	118	.24	.30	.011	.12
MATH 0080	265	.31	.40	<.001	.08
MATH 0070	420	.31	.39	<.001	.03

HS Rank

Course	*n*	r_{pb}	r_b	p_b	SD_b
ENGL 1110	863	.24	.30	<.001	.04
MATH 1110	83	.11	.14	.309	.14
MATH 0080	183	.25	.32	<.001	.09
MATH 0070	271	.27	.35	<.001	.08

Grit Scale

Course	*n*	r_{pb}	r_b	p_b	SD_b
ENGL 1110	530	.06	.07	.179	.05
MATH 1110	13	−.35	−.45	.174	.33
MATH 0080	81	.18	.22	.112	.14
MATH 0070	249	−.05	−.06	.419	.08

Note: RC = Reading Comprehension; AR = Arithmetic; EA = Elementary Algebra; CLM = College-Level Math; n = sample size; r_{pb} = point-biserial correlation; $r_{b(corrrection)}$ = biserial correlation (with correction for range restriction); p_b = p-value of biserial correlation; SD_b = standard deviation of biserial correlation.

Grit Scale. None of the correlations were no statistically significant for grit and success in English and math courses. MATH 0080 had a biserial correlation of .22 and was approaching significance at the .10 level, but this was the only result worth highlighting. Surprisingly, two of the correlations were actually negative in direction; obviously, the sample size for MATH 1110 (n = 13) was far too small to draw any conclusions from, but for MATH 0070 the sample size was nearly 250 students, and nevertheless grit was negatively related to success. Subsequent analysis did not include grit as a predictor due to these poor, inconsistent results.

Severe Error Rate Results

Table 4 shows the SER results for the four courses under analysis. Across the board, students were severely underplaced (placed into the course below the course in question when they could have earned an A or B) at higher rates than severely overplaced (placed into the course in question when they were predicted to earn an F or W there). This was especially true for MATH 1110 and MATH 0080: only 6% of students were overplaced into MATH 1110 and only 7% for MATH 0080. Given the descriptive statistics revealing just how few students even tested into these courses, it shouldn't be surprising that students were rarely overplaced there. The largest overall severe error rates using test scores alone were for ENGL 1110 and MATH 0070. Around one in three students were severely misplaced in these courses; this coheres with the correlation results that showed the two primary tests used for placement into these courses (Reading Comprehension and Arithmetic) were poor predictors of course success. By switching to HS GPA as the assessment tool and holding the proportion assigned to these levels constant, the SER was reduced by 8 percentage points in both ENGL 1110 and MATH 0070.

Simultaneously, switching to HS GPA increased the predicted success rate in these courses as well: there was an 11-percentage point gain in course success in ENGL 1110 by switching to HS GPA and a 10-percentage point gain in MATH 0070. This is a win-win for these courses: not only did switching to HS GPA reduce the rate of severe misplacement, but course success rates increased as a result of using a predictor variable that better distinguished between those likely to succeed. Using both HS GPA and test scores only slightly reduced the SER in ENGL 1110 (from 27.5% to 27.3%) and conversely lowered the predicted course success rate. For MATH 0070, using HS GPA and test scores further reduced the SER and simultaneously increased the predicted course success rate.

TABLE 4. Predicted Severe Error Rates Under Different Measures

	Measures used for Assignment		
	Test Scores	HS GPA	Test Scores + HS GPA
ENGL 1110		n = 2,243	
Severe error rate	35.0	27.5	27.3
Underplacement	19.8	16.2	16.0
Overplacement	15.2	11.2	11.3
Course success rate (≥C)	62.9	74.2	72.1
Developmental rate	49.6	49.8	49.5
MATH 1110		n = 446	
Severe error rate	25.4	32.2	24.6
Underplacement	19.6	21.8	19.0
Overplacement	5.8	10.4	5.6
Course success rate (≥C)	71.9	61.5	65.4
Developmental rate	75.4	74.9	75.2
MATH 0080		n = 2,239	
Severe error rate	17.8	17.1	15.2
Underplacement	11.3	11.2	9.9
Overplacement	6.5	5.9	5.3
Course success rate (≥C)	65.2	70.0	64.5
Developmental rate	79.7	79.7	79.6
MATH 0070		n = 1,645	
Severe error rate	31.5	23.2	21.9
Underplacement	21.4	15.5	15.5
Overplacement	10.1	7.8	6.4
Course success rate (≥C)	62.5	72.8	74.4
Developmental rate	64.0	63.8	63.9

For the higher-level math courses, there was less of a universal positive impact by incorporating HS GPA into the assessment process. Switching to only HS GPA in MATH 1110 instead of test scores increased the SER substantially and reduced the course success rate. This may be similar evidence to what Noble and Sawyer (2004) found—that standardized test scores did better in predicting performance on the high end of the achievement spectrum than did high school statistics. Using both HS GPA and test scores slightly reduced the SER compared to only using test scores, but the predicted course success rate was still lower than using test scores alone, which is a clear tradeoff.

Lastly, for MATH 0080 there were moderate reductions in the SER by switching to HS GPA; this was due in part by the fact that relatively few students were severely misplaced to begin with. Only 18% of students were severely misplaced compared to almost double that for ENGL 1110. Predicted course success increased, however, by switching to HS GPA. Using both measures reduced the SER to a mere 15%, which is especially promising, but there was a 5-percentage point drop in predicted course success.

Chapter 5. Discussion

Summary of Findings

The correlation and SER results revealed that two of the ACCUPLACER tests had low predictive validity that led to substantial rates of misplacement; high school statistics, which were quality predictors of course performance, could help to offset these tests in a multiple measures system. The Reading Comprehension and Arithmetic tests proved to be poor predictors of success in ENGL 1110 and MATH 0070 respectively, and thus the SER reductions by switching to HS GPA as an assessment tool were largest for these courses. It was also quite promising to see that switching to HS GPA or a combination of HS GPA and test scores would theoretically lessen racial and ethnic assessment gaps. This analysis saw substantially higher proportions of students of color testing into developmental English and math than their White peers. A first step in addressing these gaps is to begin incorporating HS GPA into placement decisions.

This study also verified again the near universal strength of high school transcript information in predicting college performance. Across the board, HS GPA and rank were statistically significant predictors of success in English and math courses with correlations ranging from .30 to .40 (except for HS rank and MATH 1110). In fact, I found medium-sized correlation coefficients between HS GPA and just about any college course, as well as first-term and first-year college GPA. It was encouraging to see that these local results matched what has been proven in the literature nationally.

From the SER results, using HS GPA either alone or in combination with test scores reduced the proportion of students being severely misplaced in every course except MATH 1110. For ENGL 1110, given the Reading Comprehension test's poor performance, using HS GPA alone would produce the best results.[6] For MATH 0070 and 0080, using a combination of test

---

[6] This is further confirmed by the fact that in a multiple regression predicting English course success, Reading Comprehension was not statistically significant with HS GPA in the model.

scores and HS GPA produced the best results. There were smaller gains in MATH 0080, but this was due in part, by the fact that relatively few students were severely misplaced to begin with. This cohered with the correlation results, which showed that the Elementary Algebra test, which was the only ACCUPLACER test used for placement into 0080, was the best predictor of the entire testing suite.

Lastly, the grit results showed little to no relationship with success in English and math courses. In analysis not shown here, I parsed out the grit items into the recommended second-order factors, *consistency of interest* and *perseverance of effort* (Duckworth et al., 2007) and found that *perseverance of effort* had a correlation of .08 with ENGL 1110 and was approaching significance at the .05 level. Nonetheless, this along with the MATH 0080 results above were the only grit findings that appeared to predict course performance. These inconsistent results show that grit may not be a good index of college readiness or performance although it is difficult to infer certain conclusions given such a limited sample size.

Varying Thresholds to Minimize Error Rates

Further analysis of SER estimates reveals insights into where practitioners can set placement cutoffs. Figure 3 below shows the severe error rates and predicted course success of the four courses under different percentile thresholds. These simulations use both HS GPA and test scores as assessment tools. The vertical line shows where the cutoff should be if one wanted to hold constant the proportions of students traditionally assigned to the higher and lower courses. Scott-Clayton et al. (2014) show a similar graphic as Figure 3. These graphs demonstrate that as the percentile cutoff *increases*, underplacement rates grow and overplacement rates go to zero; conversely, as the percentile cutoff *decreases*, overplacement rates grow, and underplacement rates approach zero. This should make intuitive sense—the more students one lets in, the higher the risk of overplacement, and the less students one lets in, the higher the risk of underplacement.

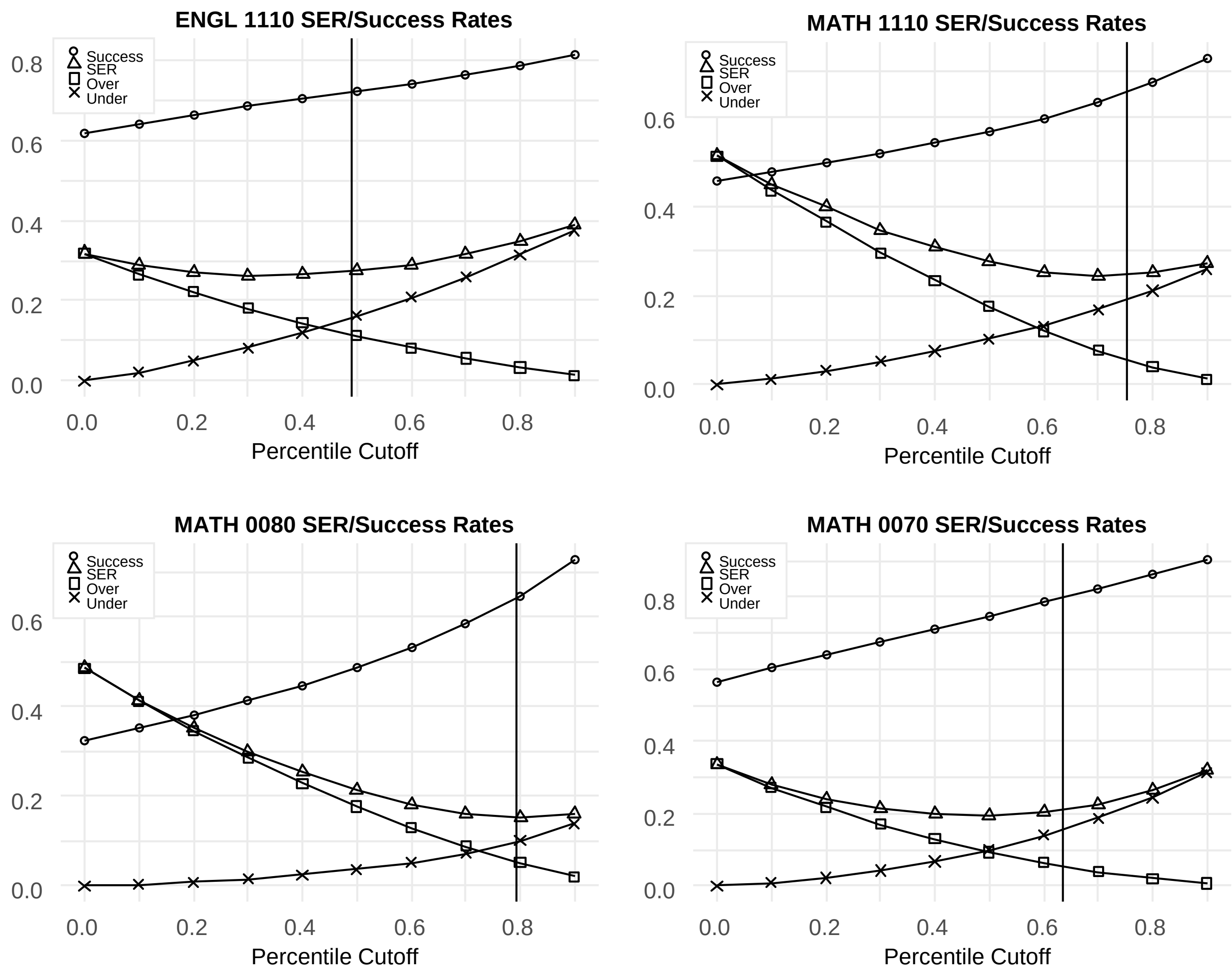

FIGURE 3. Predicted SER and Success Rates under Different Percentile Cutoffs

A second feature of these graphs is that, as Sawyer (1996) found with accuracy rates, a wide range of potential cutoffs produces very similar severe error rates. For instance, looking at the ENGL 1110 panel (top left), one can see that the percentile cutoffs between 0.2 to 0.6 produce very similar severe error rates (hovering around 27%) but vastly different severe under- and overplacement rates. This can be informative for practitioners, who may decide that underplacement is costlier an error than overplacement for their institution. If that were the case, they would set the cutoff near the lower end of that percentile cutoff band, before the under- and overplacement lines intersect.

It is unclear exactly *how* academic professionals would use the SERs to set new cutoffs for placement with multiple measures. For instance, if MCTC decided to use both test scores and HS GPA for placement into MATH 0070 and 0080, as I recommend, and they chose a percentile cutoff that minimized the severe error rates of interest, it is unknown where HS GPA and test score cutoffs should be set. A simple procedure that I propose is to find the mean HS GPA and test scores within the percentile band of interest. If for MATH 0070, 0.4 was the agreed upon cutoff, one would just find the mean HS GPA and Arithmetic and Elementary Algebra test scores between the 0.4 and 0.5 percentiles. However, cutoffs are the *lowest* score at which a student is capable of success in the course. Thus, a mean may not be the most appropriate statistic. One could instead use one or one half a standard deviation below the mean.

Another approach which does not take into account error rates is to simply use the predicted probabilities from a multiple logistic regression. A probability of 0.5 or 0.75 could be set as the cutoff, and students who are above would be placed into the higher-level course. Lastly, a more sophisticated approach is to use classification and regression trees (commonly labeled CART), a nonparametric regression procedure. This approach attempts to classify observations with similar response values given a number of predictor variables (Strobl, Malley, & Tutz, 2009). The algorithm conducts binary splits in the predictors to best group observations. For course placement, these binary decisions would act as decision rules for placing students into the appropriate courses. The multiple measures assessment project in California utilizes CART to set decision rules for placement (Newell et al., 2017). An evaluation of efforts there may indicate whether CART is the appropriate procedure for determining optimal placement cutoffs.

Limitations

The biggest single limitation of this study was the sheer volume of missing data in the sample analyzed. To compute correlations, students needed to have had both the predictor of interest and to have taken the English or math course. Many students in my dataset hadn’t taken an English or math course in the selected timeframe, which reduced the sample available for statistical analysis. Because nearly all students tested into developmental math, there were also very few observations of students taking MATH 1110 as their first math course. Further, not everyone took the entire suite of ACCUPLACER tests; this was especially acute for the CLM test, which only 13% of students took. For high school statistics and grit, the numbers were quite small as well: only 14% of students had a HS GPA on file, and only 7% took the Grit Scale. However, there have been substantial increases in collecting high school information at the college; the most recent fiscal year of enrollment data showed that close to 30% had a HS GPA on file.

Another limitation related to missing data was that I assumed that students with HS GPA were not significantly different to the general population, but from the descriptive statistics, this was not entirely true. Students with HS GPA on file were less likely to be Black or African American and tended to score marginally higher on three of the ACCUPLACER exams. Thus, one should be cautious in generalizing the SER results, which simulated using HS GPA as a placement tool, to the population of enrollees at MCTC. Additional analysis will be needed to test this assumption as the institution continues to collect more high school transcript data.

Grit was found to be unrelated to course success. However, it is unclear how the questionnaire was administered to students. Students were given the questions within the ACCUPLACER computerized suite immediately following the tests. It is unknown whether students were given any introductory material prior to taking the questionnaire, so they may not have understood its purpose. Students could have also experienced test fatigue, as the questionnaire was given at the end of placement testing. The non-significant results were still surprising, so I have chosen to present them here. This may inform others who are considering using the Grit Scale as a measure of college readiness. Although the authors of grit are hesitant to use it for individual diagnostic purposes (Duckworth & Yeager, 2015), these results add to the limited research on grit's relationship to academic performance. It is especially interesting that HS GPA, which has been known to contain such traits as persistence and effort, accounted for variance in course success where the Grit Scale did not. More research could reveal exactly which latent factors HS GPA accounts for related to college performance that questionnaires like the Grit Scale do not capture as effectively.

Lastly, in fitting logistic regression models for the SER estimates, it was difficult at times to build satisfactory models for the F or W outcome. Simply put, it proved difficult to predict who would fail or withdraw from English or math courses. Most likely, many factors contributed to students failing or withdrawing. Life circumstances not accounted for in the data could have played into these course outcomes, and consequently the models only explained a small proportion of the variance. Seeing that these models were used for estimating severe overplacement rates, one should be aware that these estimates could be slightly biased.

Future Research

Piloting new systems and evaluating their effectiveness is the logical next-step in the field of research on multiple measures assessment. Researchers could work with two-year colleges to analyze data and recommend new placement rules given high school transcript information. They could then evaluate the new system on a set of incoming students. The project at MCTC plans to conduct a small pilot to evaluate the effectiveness of the new rules that incorporate high school GPA. Based on these results, rules will be tweaked and the system scaled.

Another area of future research is to vet noncognitive assessments for diagnostic purposes. From the literature review, the LASSI and the SRI in particular appear to be two quality candidates for use in a multiple measures system. More research should be done to analyze their ability to accurately measure college readiness in English and math. Institutions and researchers could administer these assessments to students and track their course outcomes to evaluate their predictive validity. Given the significance of these results, they could then begin integrating scores on these assessments into placement decisions as well. Further research should be done to determine their viability as assessment metrics.

Lastly, as mentioned above, more analysis is needed to find the best approach for determining and assigning placements using multiple measures. I have focused on setting cutoffs or decision rules for placement, but there are many other statistical and non-statistical approaches that institutions can adopt. One helpful study on emerging practices is provided by Barnett (2017), which outlines a number of procedures such as waivers, decision bands, placement formulas, decision rules, and directed self-placement. As to which method is best is a question that should be explored in future studies.

Conclusion

First-year academic performance in college is critical for future retention, transfer, and graduation (Allen et al., 2008). It is imperative that two-year institutions make sound decisions around course placement to ensure students are in the most appropriate courses for skill development and success. The first step in making such quality decisions is a systematic review of current practice and analysis of other viable predictors beyond traditional placement exams. This thesis has taken on that task for one two-year institution that serves a highly diverse student population. I hope that the findings presented here can be used at other community colleges to aid in evaluating assessment policies and practices, with the ultimate goal of benefiting students and their educational aspirations.

References

Allen, J., Robbins, S. B., Casillas, A., & Oh, I.-S. (2008). Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness. *Research in Higher Education*, *49*(7), 647–664. Retrieved from http://www.jstor.org/stable/25704590

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *The Journal of Higher Education*, *77*(5), 886–924. https://doi.org/10.1353/jhe.2006.0037

Bailey, T. (2008). Challenge and opportunity: Rethinking the role and function of developmental education in community college. *New Directions for Community Colleges*, *145*(161), 81. https://doi.org/10.1002/cc

Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, *29*(2), 255–270. https://doi.org/10.1016/j.econedurev.2009.09.002

Barnett, E. A. (2017). *College placement strategies evolving considerations and practices*. Center for the Analysis of Postsecondary Readiness. Retrieved from https://ccrc.tc.columbia.edu/media/k2/attachments/college-placement-strategies-evolving-considerations-practices.pdf

Belfield, C. R., & Crosta, P. M. (2012). *Predicting success in college: The importance of placement tests and high cchool transcripts*. (Working Paper No. 42). Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/media/k2/attachments/predicting-success-placement-tests-transcripts.pdf

Burdman, P. (2012). Where to begin? The evolving role of placement exams for students starting college. *Jobs for the Future*, (August), 1–38. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED537265

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

https://doi.org/10.1038/141613a0

College Board. (2004). *ACCUPLACER coordinator's guide*. New York: College Board. Retrieved from http://www.olc.edu/~cdelong/ACCUPLACER/CoordinatorGuide.pdf

College Board. (2013). *Improving the effectiveness of placement tests*. New York: College Board. Retrieved from https://accuplacer.collegeboard.org/educator/multiple-weighted-measures

College Board. (2016). *ACCUPLACER program manual*. New York: College Board. Retrieved from https://secure-media.collegeboard.org/digitalServices/pdf/accuplacer/accuplacer-program-manual.pdf

Conley, D. T. (2007). *Redefining college readiness*. Eugene, OR: Educational Policy Improvement Center. https://doi.org/10.1002/he.321

Conley, D. T. (2012). *Definition of college and career readiness*. Eugene, OR: Educational Policy Improvement Center. Retrieved from http://www.epiconline.org/download/36997/

Cortina, J. M., & Nouri, H. (2000). *Quantitative applications in the social sciences: Effect size for ANOVA designs*. SAGE Publications, Inc. https://doi.org/10.4135/9781412984010

Crede, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science*, *3*(6), 425–453. https://doi.org/10.1111/j.1745-6924.2008.00089.x

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (Grit-S). *Journal of Personality Assessment*, *91*(2), 166–174. https://doi.org/10.1080/00223890802634290

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–

251. https://doi.org/10.3102/0013189X15584327

Fergus, M. (2016). *Getting prepared*. Minnesota Office of Higher Education. Retrieved from https://www.ohe.state.mn.us/pdf/GettingPrepared2016.pdf

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Los Angeles: SAGE.

Fields, R., & Parsad, B. (2012). *Tests and cut scores used for student placement in postsecondary education: Fall 2011*. National Assessment Governing Board. Retrieved from https://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/surveys/test-and-cut-scores-used-for-student-placement-in-postsecondary-education-fall-2011.pdf

Fox, J. (2016). *Applied regression analysis and generalized linear models* (3rd ed.). Los Angeles: SAGE.

Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High school record vs. standardized tests as indicators of four-year college outcomes*. (Research & Occasional Paper Series 6.07). University of California, Berkeley. Retrieved from http://www.cshe.berkeley.edu/sites/default/files/shared/publications/docs/ROPS.GEISER._SAT_6.13.07.pdf

Hughes, K. L., & Scott-Clayton, J. (2011). Assessing developmental assessment in community colleges. *Community College Review*, *39*(4), 327–351. https://doi.org/10.1177/0091552111426898

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average*. New York: College Board. Retrieved from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2008-5-validity-sat-predicting-first-year-college-grade-point-average.pdf

Komarraju, M., Ramsey, A., & Rinella, V. (2013). Cognitive and non-cognitive predictors of college readiness and performance: Role of academic discipline. *Learning and Individual Differences*, *24*, 103–109. https://doi.org/10.1016/j.lindif.2012.12.007

Le, H., Casillas, A., Robbins, S. B., & Langley, R. (2005). Motivational and skills, social, and self-management predictors of college outcomes: Constructing the Student Readiness Inventory. *Educational and Psychological Measurement*, *65*(3), 482–508. https://doi.org/10.1177/0013164404272493

Long, M. C., Iatarola, P., & Conger, D. (2009). Explaining gaps in readiness for college-level math: The role of high school courses. *Education Finance and Policy*, *4*(1), 1–33. https://doi.org/10.1162/edfp.2009.4.1.1

Maruyama, G. (2012). Assessing college readiness: Should we be satisfied with ACT or other threshold scores? *Educational Researcher*, *41*(7), 252–261. https://doi.org/10.3102/0013189X12455095

Mattern, K. D., & Packman, S. (2009). *Predictive validity of ACCUPLACER scores for course placement: A meta-analysis*. New York: College Board. Retrieved from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2009-2-predictive-validity-accuplacer-scores-course-placement.pdf

Melguizo, T., Kosiewicz, H., Prather, G., & Bos, J. (2014). How are community college students assessed and placed in developmental math? Grounding our understanding in reality. *Journal of Higher Education*, *85*(5), 691–722. https://doi.org/10.1353/jhe.2014.0025

Minneapolis Community and Technical College. (2017). About us. Retrieved March 26, 2017, from http://www.minneapolis.edu/about-us

Morgan, D., & Michaelides, M. P. (2005). *Setting cut scores for college placement*. New York: College Board. Retrieved from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-

2005-9-setting-cut-scores-college-placement.pdf

Newell, M., Hayward, C., Lamoree, D., Hetts, J., Fagioli, L., Bahr, P., … Sorey, K. (2017). *Multiple measures—high school variables model summary, phase II*. RP Group. Retrieved from http://rpgroup.org/Portals/0/Documents/Projects/MultipleMeasures/DecisionRulesandAnalysisCode/Statewide-Decision-Rules-5_18_16_1.pdf

Ngo, F., & Kwon, W. W. (2014). Using multiple measures to make math placement decisions: Implications for access and success in community colleges. *Research in Higher Education*, 442–470. https://doi.org/10.1007/s11162-014-9352-9

Noble, J. P., & Sawyer, R. (2004). Is high school GPA better than admission test scores for predicting academic success in college? *College and University Journal*, *79*(4), 17–22. Retrieved from https://eric.ed.gov/?id=EJ739081

Parsad, B., Lewis, L., & Greene, B. (2003). *Remedial education at degree-granting postsecondary institutions in fall 2000*. Jessup, MD: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs2004/2004010.pdf

Peterson, C. H., Casillas, A., & Robbins, S. B. (2006). The Student Readiness Inventory and the Big Five: Examining social desirability and college academic performance. *Personality and Individual Differences*, *41*(4), 663–673. https://doi.org/10.1016/j.paid.2006.03.006

Porchea, S. F., Allen, J., Robbins, S. B., & Phelps, R. P. (2016). Predictors of long-term enrollment and degree outcomes for community college students : Integrating academic, psychosocial, socio-demographic, and situational factors. *The Journal of Higher Education*, *81*(6), 680–708. Retrieved from http://www.jstor.org/stable/40929572

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261–88. https://doi.org/10.1037/0033-2909.130.2.261

RP Group. (2017). Multiple measures assessment project. Retrieved April 23, 2017, from http://rpgroup.org/All-Projects/ctl/ArticleView/mid/1686/articleId/118/Multiple-Measures-Assessment-Project-MMAP

Sawyer, R. (1996). Decision theory models for validating course placement. *Journal of Educational Measurement*, *33*(3), 271–290. https://doi.org/10.1111/j.1745-3984.1996.tb00493.x

Sawyer, R. (2007). Indicators of usefulness of test scores. *Applied Measurement in Education*, *20*(3), 255–271. https://doi.org/10.1080/08957340701431245

Sawyer, R. (2013). Beyond correlations: Usefulness of high school GPA and test scores in making college admissions decisions. *Applied Measurement in Education*, *26*(2), 89–112. https://doi.org/10.1080/08957347.2013.765433

Scott-Clayton, J. (2012). *Do high-stakes placement exams predict college success?* (Working Paper No. 41). New York: Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/media/k2/attachments/high-stakes-predict-success.pdf

Scott-Clayton, J., Crosta, P. M., & Belfield, C. R. (2014). Improving the targeting of treatment: Evidence from college remediation. *Educational Evaluation and Policy Analysis*, *36*(3), 371–393. https://doi.org/10.3102/0162373713517935

Strayhorn, T. (2014). What role does grit play in the academic success of Black male collegians at predominantly White institutions. *Journal Of African American Studies*, *18*(1), 1–10. https://doi.org/10.1007/s12111-012-9243-0

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323–348. https://doi.org/10.1037/a0016973

Venezia, A., Bracco, K. R., & Nodine, T. (2010). *One-shot deal? Students' perceptions of assessment and course placement in California's community colleges*. San Francisco, CA:

WestEd. Retrieved from https://www.wested.org/online_pubs/oneshotdeal.pdf

Wiberg, M., & Sundström, A. (2009). A comparison of two approaches to correction of restriction of range in correlation analysis. *Practical Assessment, Research & Evaluation*, *14*(5), 1–9. Retrieved from http://www.pareonline.net/getvn.asp?v=14&n=5

Wolters, C. A., & Hussain, M. (2015). Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacognition and Learning*, *10*(3), 293–311. https://doi.org/10.1007/s11409-014-9128-9