THE TEXAS SUCCESS INITIATIVE ASSESSMENT AND THE TESTS OF ADULT BASIC EDUCATION: A COMPARISON OF STUDENT PERFORMANCE AND PLACEMENT

by

Denise Annette Guckert, B.A., M.A.

A dissertation submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Doctor of Education
with a Major in Developmental Education
December 2021

Committee Members:

Russ Hodges, Chair

Eric J. Paulson

Robert Reardon

Taylor Acee

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

ACKNOWLEDGEMENTS

There are a few people who have helped me more than they probably wanted to and more than I deserved. First, I must acknowledge Dr. Russ Hodges. He always has believed in me and encouraged me to finish. Over the many, many years I have worked on this, I have grown to know him as a truly wonderful scholar, a tireless teacher, and a truly gracious and generous person.

I want to thank my father, who is always there if I need him. He inspired me to get a doctorate because he earned his when I was six years old! I remember him taking off on his bicycle to Texas A & M every morning while my mother kept our one car to haul her 3 young children around.

I want to thank Juan Villedas, my truly dear partner who has taken care of everything for me as I have worked towards the end of my studies. I also thank my friend Melissa Sadler-Nitu for her help along the way.

Finally, I want to acknowledge my committee members: Dr. Hodges, my chair, Dr. Paulson, my philosophical prodder, Dr. Acee, my questioner, and Dr. Reardon, my statistics guru who go me through my results. These professors patiently waited through the many years it took me to complete my comps and dissertation and provided me with insightful input.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

ABSTRACT

This study explores the relationship between scores and educational placement levels on two academic mathematics assessment systems: The Tests of Adult Basic Education 11/12 mathematics (TABE-M) and the Texas Success Initiative Assessment mathematics (TSIA-M). The data set consisted of scores, placement levels, and demographic information for 152 students who enrolled in a College Success Academy at a small community college in Texas. With no mathematics instruction between the closely administered assessments, it was possible to correlate student scores and placement levels on the assessments to investigate the convergent validity of the scores and consistency of classification for the placement levels. Fischer's r-to-z transformation tests and calculation of confidence intervals were used to determine if correlations between student performance varied across student demographic characteristics. Findings of convergent validity support the use of TSIA-M for its use in identifying students with mathematic skills below secondary levels but did not support the use of the TSIA-M’s Adult Basic Education (ABE) Diagnostic for accurate student placements into developmental education or ABE instruction. Study results can assist advisors with decisions about placement into various types of college readiness instruction. The TSIA-M ABE Diagnostic overplaced 80% of by up to two ABE levels, making it highly probable that students with level 5 placements were incorrectly placed into DE courses without basic skills support or into corequisite DE and that students with level 3 placements were not advised of options for referral to adult education services.

I. INTRODUCTION

Almost all college journeys begin with course placement that is often highly dependent on high-stakes, standardized testing (Hughes & Scott-Clayton, 2011). Institutions of higher education (IHEs) across the United States use SAT or ACT college admissions test scores along with other factors in a competitive selection process to admit beginning year students (Atkinson & Geiser, 2009). In contrast, community college systems throughout the United States are primarily open access and rarely deny entrance to a student (Desai, 2012). Though community colleges are typically open access, 100% of 970 public 2-year institutions nationwide reported using high stakes test scores for placement purposes in at least one academic area. Colleges use these assessment scores alone or in conjunction with other measures, such as recent high school grades, for placement into college-level versus development education coursework (Fields & Parsad, 2012).

As do many states across the nation, Texas uses scores from a college placement test, which in Texas is the mandated Texas Success Initiative Assessment (TSIA), to determine college readiness and inform course placement (19 Tex. Admin. Code §4.56). However, colleges may also use TSIA results to inform decisions to refer college applicants to federally funded adult basic education services rather than to place the student in college developmental education coursework or interventions (The Higher Education Coordinating Board [THECB], 2015). Although the THECB provides empirical evidence in support of the TSIA's college readiness cut-off scores for placement into college coursework (Cui & Bay, 2017), no research exists that examines the validity of the TSIA ABE Diagnostic assessment that places students into various levels of developmental education or adult education services.

Background of the Problem

Results from studies that focus on the validity of the placement tests have called into question the use of college placement test scores as a sole measure for determining academic college readiness (Bracco et al., 2014; Hughes & Scott-Clayton, 2011; Scott-Clayton, 2012). In turn, many states are changing college placement testing and policies. Revisions in state policy include the development of college placement tests that align to state- or state-consortia-developed college and career standards and the stipulation that colleges use multiple measures, including high school grade point average (HSGPA), to aid in placement decisions. Some states allow students in the K–12 system who proceed directly to postsecondary to use standardized academic tests taken in high school for placement purposes (Barnett et al., 2018; Education Commission of the States, 2017; Schak et al., 2017). Researchers who examine the predictive ability of HSGPA and college admissions and placement test scores repeatedly report that combining HSGPA with test scores results in better placement and admissions decisions than does using either measure used alone (Mattern & Packman, 2009; Noble & Sawyer, 2013; Westrick & Allen, 2014; Westrick et al., 2015). Even so, as of 2016, only 51% of community colleges surveyed reported that they were using any information other than college placement test scores for placing students into reading or writing college-level or developmental (DE) instruction; for mathematics, the percentage was 57% (Rutschow & Mayer, 2018).

In Texas, scores on the TSIA and TSIA diagnostic assessments determine student placement into traditional college-level or corequisite college courses and inform decisions on DE placement or referral to federally funded adult basic education services. Notably, scores from both the TSIA diagnostic assessments and the TABE 11/12 are aligned to a common scale of federally defined adult basic skills proficiency levels (THECB, 2014). This scale, developed by the U.S. Department of Education's National Reporting System for Adult Education (NRS), consists of six adult basic skills performance levels known as NRS Educational Functional Levels (NRS levels) (NRS, 2021).

Statement of the Problem

For students in Texas who lack current or complete high school records, college placement testing plays an elevated role as students must demonstrate their academic proficiency based on college placement or admissions test scores. Adults who earn a high school equivalency certificate by examination rather than a traditional high school diploma lack complete high school transcripts that college advisors could use to make college admissions or placement decisions. Adults who delay college entrance until several years after earning a high school diploma or high school equivalency certificate may also have difficulty in providing these transcripts. Therefore, adults in Texas with incomplete or outdated high school records experience the heightened importance of TSIA scores and placements.

The Texas legislature requires that as of the 2020–2021 academic year, 100% of college applicants who are college-ready or who place above ABE Diagnostic Levels 1–4 (analogous to NRS levels) on the TSIA be placed into mainstreamed college corequisite courses (19 Tex. Admin. Code §4.62[8]). For students who place into ABE Levels 1 or 2 on TSIA diagnostic assessments, THECB strongly recommends referral to federally funded adult basic education (FF-ABE) services. Because NRS levels 1 and 2 correspond to basic skills below fourth-grade levels, THECB has worked to partner with the Texas Workforce Commission, which administers FF-ABE services, so that students at these lower literacy levels could benefit from free FF-ABE services until they are ready to transition to postsecondary (THECB, 2008). THECB's recommendation for students who place into ABE Levels 3 or 4 (fifth- to eighth-grade skill levels) is enrollment in a stand-alone DE course with concurrent enrollment in a basic skills support intervention (THECB, 2015). However, colleges have discretion for supporting students below ABE levels 5 or 6; they may refer them to college DE interventions or coursework or to FF-ABE services that may or may not be located within the college (Daugherty et al., 2019a).

Throughout my 30 years of experience working as an instructor and administrator in various FF-ABE programs in Texas, I have partnered with and worked directly for community colleges that house FF-ABE programming. Due to closer alignment of FF-ABE and postsecondary policies designed to help students at the lowest literacy levels succeed in college (THECB, 2008), some community colleges consolidated resources to develop college transitions programs that served the needs of FF-ABE students who were transitioning to postsecondary as well as students whom the TSIA placed below DE academic levels. Therefore, I began to see students in the FF-ABE program who were referred there by their placements from the TSIA ABE Diagnostic. In this capacity, I have observed and also collected anecdotal evidence from FF-ABE practitioners that student placements on their TSIA ABE Diagnostic and their FF-ABE TABE 11/12 test were often quite disparate (K. Dowdy, D. Johnston, S. Smith, & L. Webb, personal communication, February 1, 2019; J. Hayes, M. Sadler-Nitu & G. Rose, personal communication, September 27, 2019). With the adoption of the TABE 11/12 beginning in the 2018 academic year, FF-ABE practitioners reported seeing a trend with the TSIA ABE Diagnostic often placing students into higher NRS levels than the TABE 11/12 (J. Hayes, M. Sadler-Nitu, and G. Rose, personal communication, September 27, 2019). Because both the TSIA ABE Diagnostics and the TABE 11/12 were specifically developed to place students into the same federal measurement scale known as the NRS (DRC, 2017; THECB, 2014), NRS placement levels should not have varied widely. I became concerned about the misplacement of students as determined by the TSIA ABE Diagnostics and desired to see if there was a statistically significant difference in student scores and placements between the two assessments.

Purpose of the Study

The naturally occurring closely timed administration of the TSIA subtests and their adult basic education diagnostics and the TABE 11/12 subtests to college applicants who were referred to and enrolled in federally funded adult education programs made it possible to collect and analyze adult learners' performance on the two tests. The purpose of this study, therefore, was to explore the relationships between student performance on two closely administered tests that generated test scores along with NRS level placements.

Per researchers Carlson and Herdman (2010), the construct of a test is an underlying ability or skill of the test-taker that a test purports to measure that "reflects the extent to which two measures capture a common construct" (p. 19). Convergence evidence adds to the validation argument that a test's content measures what it was designed to measure. To determine convergent validity, researchers calculate the correlation between an unproven and a well-established assessment. The stronger a correlation is, the more likely the two tests measure the same construct. A moderate to strong correlation supports convergent validity, which in turn supports the construct validity of the unproven assessment.

Findings of statistically moderate or strong correlations between scores on the TSIA and TABE 11/12 subtests would provide convergent validity evidence that these instruments were measuring similar constructs and would support the intended use of the TSIA subtests to identify students with academic skills below high school levels who needed to be administered TSIA ABE Diagnostics for placement purposes. A moderate to strong correlation between student TSIA ABE Diagnostics and TABE 11/12 subtest placement levels would provide convergent validity evidence that these instruments measured similar constructs and build support that the TSIA ABE Diagnostics measured constructs similar to the TABE 11/12 subtests.

Consistency of classification analysis is one method that test developers use to support the intended use of test scores for student placement. In this study, I could compare TSIA ABE Diagnostic placement levels to TABE 11/12 subtest placement levels to determine the consistency of classification of the TSIA ABE Diagnostics into the same levels as the TABE-M subtests. According to DRC (2017), to establish adequate consistency of classification to support the use of a test for student placement, "hit rates of at least 50% are desirable; higher rates are the most desirable" (p. 78). Therefore, if the consistency of classification rate for the TSIA ABE Diagnostics and the TABE 11/12 tests are at least 50%, the TSIA ABE Diagnostic may be considered adequate for its use for student placement into adult basic skills levels.

Assumptions

For this investigation, I assumed that the student demographic and assessment information from the college's educational records were accurate. I also assumed that digital software such as SPSS, spreadsheet software, and online calculators functioned correctly. I further assumed that the TABE 11/12—one of only two assessments that is federally approved for use in determining mathematics placement into all NRS levels—is one of the most reliable and valid assessments of adult mathematics proficiency at primary and secondary levels (Tests Determined To Be Suitable for Use in the National Reporting System for Adult Education, 2020). According to the U.S. Department of Education, "A test must meet all applicable and feasible standards for test construction and validity provided in the 1999 edition of the *Standards for Educational and Psychological Testing"* (Adult Education and Family Literacy Act, 2011). See Chapter 3 for further evidence of the TABE 11/12's validity and reliability.

Scope

My original intent for this study was to compare student performance on all sections—mathematics, mathematics, reading, and language—that comprise the TSIA, the TSIA ABE Diagnostic, and the TABE 11/12. However, data collected at the study site did not include sufficient student records with reading and language assessments. Therefore, the scope of the study was limited to the analysis of the student performance on the mathematics sections only on the TSIA, TSIA ABE Diagnostic, and TABE 11/12.

The student population was limited to students who met the following criteria: (a) applied to a small Texas community college; (b) completed the mathematics sections of the TSIA (TSIA-M) and the TSIA adult basic education diagnostic (TSIA-M ABE Diagnostic) and received scores from both assessments (c) completed TABE 11/12 mathematics (TABE-M) testing within 6 weeks of their TSIA testing; and (d) received no academic instruction between the TSIA-M and TABE-M test administrations.

Research Questions and Null Hypotheses

The purpose of this investigation is to compare student performance on closely administered TSIA-M, TSIA-ABE Diagnostic, and TABE-M assessments by analyzing the relationships between student scores and placement levels on these instruments. Below are the research questions I formulated:

- Research Question 1: What is the relationship between student scores on closely administered TSIA-M and TABE-M assessments?
- Research Question 2: What is the relationship between student NRS placements on closely administered mathematics TSIA-M ABE Diagnostic and TABE-M assessments?
- Research Question 3: Does controlling for age, gender, race, or ethnicity change the relationships between student scores on closely administered mathematics TSIA-M and TABE-M assessments?
- Research Question 4: Does controlling for age, gender, race, or ethnicity change the relationships between student NRS placements on closely administered mathematics TSIA-M ABE Diagnostic and TABE-M assessments?

To answer the above research questions, I tested the following null hypotheses:

- Null Hypothesis 1: There is no statistically significant relationship between student scores on closely administered mathematics TSIA-M and TABE-M assessments.
- Null Hypothesis 2: There is no statistically significant relationship between student NRS placements from closely administered mathematics TSIA-M ABE Diagnostic and TABE-M assessments.
- Null Hypothesis 3: There are no statistically significant differences in the relationships between student scores on closely administered mathematics TSIA-M and TABE-M assessments after controlling for age, gender, race, or ethnicity.
- Null Hypothesis 4: There are no statistically significant differences in the relationships between student NRS placements on closely administered mathematics TSIA-M ABE Diagnostic and TABE 11/12 assessments after controlling for age, gender, race, or ethnicity.

Definition of Terms

Adult basic education (ABE): "Activities and instruction are provided across a continuum, from pre- and basic literacy through elementary levels, culminating with competencies equivalent to the eighth-grade level" (Texas Workforce Commission, 2017, p. 38).

Adult basic education levels: TSIA ABE Diagnostic tests scores ranging from 1 to 6 that align with the National Reporting System for Adult Education Educational Functioning Levels (Daugherty et al., 2019a).

Adult secondary education (ASE): "Activities and instruction are comparable to the competencies developed in secondary high school and college developmental education" (Texas Workforce Commission, 2017, p. 38).

Basic adult skills education (BASE): In Texas public postsecondary institutions, these are "students who are directed by ABE Diagnostic with Levels 3–4 on the TSIA ABE Diagnostic" (Morales-Vale & Daniels, 2015, p. 18).

Basic adult skills education non-course competency-based developmental education option (BASE NCBO): Basic academic skills interventions that occur concurrently with lower-level stand-alone DE classes. "BASE NCBOs aligned with DE courses are considered HS knowledge and skill levels" (Morales-Vale & Daniels, 2015, p. 26).

College and Career Readiness Standards for Adult Education (CCR-AE): "A set of college and career readiness standards that reflect the content most relevant to preparing adult students for success in colleges, technical training programs, work, and citizenship—in the areas of English language arts/literacy and mathematics" (Literacy Information and Communication System, n.d., para. 1). "The work is the result of a nine-month process that examined the Common Core State Standards from the perspective of adult education" (Office of Career and Technical Adult Education, n.d., para. 2).

Common Core State Standards (CCSS): Often referred to as the Common Core, CCSS are a "set of clear college- and career-ready standards for kindergarten through 12th grade in English language arts/literacy and mathematics. Today, 41 states and the District of Columbia have voluntarily adopted and are working to implement the standards" (CCCS Initiative, 2018, para. 2).

Concurrent validity: Concurrent validity is the "correlation between a measure and a standard regarded as a representative of the construct under consideration . . . If the measure is correlated with an assessment in the same general time frame, this is termed concurrent validity" (West & Beckman, 2018, p. 357).

Convergent validity: According to researchers Carlson and Herdman (2010), convergent validity is determined through correlation. "Convergent validity reflects the extent to which two measures capture a common construct" (p. 18). "It is a form of evidence used to judge the construct validity of a measure" (pp. 19–20).

Developmental coursework or intervention: "Non-degree credit coursework and/or activity designed to address a student's strengths and needs in the areas of reading, writing, integrated reading and writing (IRW), mathematics, and student success" (19 Tex. Admin. Code §4.53[9]).

Developmental education (DE): "Pre-college, non-degree credit courses, interventions, tutorials, laboratories, and other means of assistance that are included in a plan to ensure the success of a student in performing entry-level academic coursework" (19 Tex. Admin. Code §4.53[10]). According to THECB, "DE, guided by the National Reporting System for Adult Education Educational Functioning Level Descriptors . . . is now defined in Texas as knowledge and skill levels at ninth through twelfth grade (levels 5–6). ABE is defined as knowledge and skill levels at first through eighth grade (levels 1–4)" (THECB, 2014, p. 33).

Federal-funded adult basic education (FF-ABE): Services that use federal funds allocated by their states to provide programs that "assist adults to become literate and obtain the knowledge and skills for employment and economic self-sufficiency" (Uvin, 2015, p. 2).

General education development certificate (GED): A higher secondary certificate, accepted by 97% of institutions in the United States in lieu of a high school diploma, that is earned by passing a series of tests developed by the GED Testing Service (GED Testing Service, 2019a, 2019b).

Recognized equivalent of a high school diploma (HSE): "The following are the equivalent of a high school diploma – (1) A General Education Development Certificate (GED); (2) A State certificate received by a student after the student has passed a State-authorized examination that the State recognizes as the equivalent of a high school diploma" (Higher Education Act of 1965, as Amended, 2016).

National Reporting System for Adult Education (NRS): The outcome-based "accountability system for the federally funded, State[sic]-administered adult education program" (NRS, 2021, p. 1).

NRS educational functioning levels (NRS levels): "The ABE, ASE, and ESL literacy levels, as provided in the Guidelines, that describe a set of skills and competencies that students demonstrate in the NRS skill areas" (Adult Education and Family Literacy Act, 2018, p. 127).

Non-course competency-based developmental education option (NCBO): In Texas, these are instructional interventions for students below college readiness levels "that use learning approaches designed to address a student's identified weaknesses and effectively and efficiently prepare the student for college-level work. These interventions must be overseen by an instructor of record, must not fit traditional course frameworks, and cannot include advising or learning support activities already connected to a traditional course" (19 Tex. Admin. Code §4.53[18]).

Predictive validity: *"*A type of criterion validity, which concerns the correlation between a measure and a standard regarded as a representative of the construct under consideration. If the measure is correlated with a future assessment, this is termed predictive validity" (West & Beckman, 2018, p. 357).

Tests for Adult Basic Education 11/12 (TABE 11/12): "The most comprehensive and reliable academic assessment product in adult basic education" (Data Recognition Corporation [DRC], 2019, para. 1).

Texas Higher Education Coordinating Board (THECB): The Texas governmental agency that is the "highest authority in the state in matters of public higher education" (THECB, 2019, para. 2).

Texas Success Initiative (TSIA): Texas legislation that mandates college readiness assessment in mathematics, reading, and writing for all entering students who do not meet specified exemptions. The legislation stipulates that non-exempt students who are not college-ready "be provided with a plan for academic success which may include corequisite or other developmental education courses/interventions" (THECB, 2017, p. 1).

Texas Success Initiative Assessment (TSIA): "assessment instrument . . . for use by institutions of higher education for assessing a student's readiness to enroll in an entry-level freshman course" (19 Tex. Admin. Code §453 [23]).

Texas College and Career Readiness Standards (TXCCRS): Curriculum standards that "address what students must know and be able to do to succeed in entry-level college courses offered at Texas public community/technical colleges and universities" (THECB, n.d.-d, para. 1).

Texas Workforce Commission (TWC): The Texas governmental agency "charged with overseeing and providing workforce development services to employers and job seekers of Texas" (Texas Workforce Commission, 2021, para. 1).

Workforce Innovation & Opportunities Act (WIOA): Federal legislation "designed to strengthen and improve our nation's public workforce system and help get Americans, including youth and those with significant barriers to employment, into high-quality jobs and careers and help employers hire and retain skilled workers" (U.S. Department of Labor, n.d., para. 1).

Worldview

My worldview, which served as a broad approach to my investigation, has been that of the pragmatic researcher. The word pragmatism comes from the Greek root pragma, meaning action. Pragmatic philosophy is, therefore, based on actions and one's position to them. The primacy of experience in forming beliefs and the importance of consequences on truth underly the pragmatic worldview (Goldkuhl, 2012). Pragmatist researchers are problem-centered, pluralistic, grounded in real-world practice, and interested in the consequences of actions (Creswell & Plano Clark, 2007). They share aspirations to create “a philosophy that fully respects the modern scientific worldview without thereby losing contact with the world of human practice” (Talisse & Aikin, 2011, p. 9).

Neither dogmatic in nature nor skeptics, pragmatists view knowledge as meaning that we as humans construct through our actions within the world (Thayer, 1966). Pragmatists reject the idea of immobile, absolute truths about knowledge. Instead, pragmatists maintain that truths reflect our experiences in an ever-changing world and are therefore malleable and fallible (Gutek, 2012). Moreover, truth must be evaluated not just on past experiences and present realities; the consequences of actions based on that truth must also be evaluated (Bleasby, 2011).

The well-known pragmatic philosopher Dewey avoided answering the age-old philosophical question of “What is truth?” Instead, he searched for warrantable assertions, which he believed was “a better way of capturing the function of both knowledge and truth insofar as both are goals of inquiry” (Capps, 2019, para. 23). In his theory of inquiry, Dewey (1938) outlined an inquiry process with three distinct stages: the indeterminate situation, the problematic situation, and the determinate situation (Brown, 2012). The indeterminate situation began with doubt, which “is uneasy; it is tension that finds expression and outlet in the processes of inquiry” (Dewey, 1938, p. 8). In the problematic situation, speculation led to the formalization of null hypotheses and the development of actions that might resolve the doubt. The determinate situation occurred when the hypothesis testing has concluded, and a warranted assertion based on findings could be stated. A key point in Dewey’s theory of inquiry is that “it is the convergent and cumulative effect of continued inquiry that defines knowledge in its meaning. The attainment of settled meaning is a progressive matter; there is no belief so settled as not to be exposed to further inquiry" (Dewey, 1938, pp. 8–9).

In the case of this study, my indeterminate situation began with my realization of an apparent mismatch in students' NRS level placements between students' TSIA ABE Diagnostic tests and the TABE 11/12. Because the TABE 11/12 is an established assessment of primary and secondary level mathematics skills and placement into NRS levels, I began to question the accuracy of the TSIA ABE Diagnostic placements. I then entered Dewey's problematic situation phase, in which I defined the problem and considered plausible explanations. If anecdotal evidence that the TSIA ABE Diagnostics tended to place students at higher NRS levels than the TABE 11/12 were true, I initially hypothesized that this may have been due to a lack of adequate piloting of diagnostic components of the TSIA with students who had skills significantly lower than college-ready. Alternatively, I hypothesized that the differences in student placement levels could be due to the cut scores used by the TSIA ABE Diagnostic to place students into the NRS levels.

However, my plausible explanations on why placements levels were not aligning well were based only on anecdotal evidence. With no technical report from the test publisher available, I needed to first determine if there was indeed a significant difference between the placement levels from the TSIA ABE Diagnostics and the TABE 11/12. If there was a significant difference, I wondered if any discernable pattern in these differences would help with adult education transition to postsecondary and college advisor placement for students who were below ninth-grade levels.

It was at this point in my inquiry that my thought process crystallized, which enabled me to develop my research questions, determine my research lens, and begin my review of relevant research. However, I was still within Dewey's (1938) problematic situation of the inquiry process. My investigation of relevant literature led me to an iterative process of discovery and my development of a conceptual framework for test validation, which in turn framed further investigation leading to hypothesis testing based on my research design. Once results were in hand, I needed to reflect, which is itself a process. According to Dewey (1910), "reflection involves not simply a sequence of ideas, but a *consequence*—a consecutive ordering in such a way that each determines the next as its proper outcome" (pp. 2–3). Finally, my inquiry transitioned into the final phase—the determinant situation—as I analyzed my findings within the context of other relevant research and suggested possible consequences of those findings for creating change in future actions and in stimulating further investigation.

Dewey wrote that "It is therefore not the origin of a concept, it is its application which becomes the criterion of its value; and here we have the whole of pragmatism in embryo" (Dewey, 1931, p. 29). In other words, Dewey viewed pragmatism as the practical application of knowledge in solving real problems in everyday experiences. It is my intent for the knowledge I have discovered through this investigation to prove useful and practical information for advisors involved in the placement of students into FF-ABE adult education or postsecondary developmental instruction services as well as to FF-ABE practitioners who support adults as they begin their postsecondary journeys.

Theoretical Framework

The original intent of my investigation was to explore relationships between the TSIA and the TABE 11/12 to discern information that could help adult basic education and postsecondary practitioners make more informed decisions regarding student placement. However, results from research that analyzes measurement scores or the intended use of measurement scores to one or more external variables are sources of evidence in the test validation process (American Educational Research Association [AERA] et al., 2014). Therefore, researching the theory of test validity and how concepts of test validity might frame my research, methodology, results, and discussion emerged as a necessary component of my investigation.

The Development of Test Validity Theory

Discussion of test validity began in the early 1900s in the context of intelligence testing. Buckingham (1921) defined test validity as "the extent to which they [tests] measure what they purport to measure" (p. 274) and asserted that validity was even more important than reliability. Buckingham foreshadowed many current tenants of test validity, including test linkages, the right for the public to be aware of a test's psychometric properties, the need to establish validity through correlational investigation. Kelley (1927) also predicted many future social and philosophical issues that would surface around concepts of test validity, such as the assumption of unidimensional traits, content representation that advantages or disadvantages certain test-takers, and score interpretation.

In the mid-1900s came the recognition that test validity must include not only content and criterion-referenced evidence but also construct validity, which "involves first making predictions based on theory then gathering data to confirm those" (American Psychological Association [APA], 1954, as cited in Shepard, 1993, p. 416). Cronbach and Meehle (1955) proposed that construct validity required the construction of a nomological network to link evidence to support theoretical constructs and a validation process that justified the instrument's validity.

Messick (1989) proposed a "new unified concept of validity…a more comprehensive theory of construct validity addressing both score meaning and social values in both test interpretation and test use" (Messick, 1994, p. 5). In his unified theory of construct validity, Messick included six aspects of construct validity that must be considered in the validation process. Although Messick's theory is considered the most well-known theory of validity (Royal, 2017; Zumbo & Chan, 2014), not all theorists agree with the Messick's (1989) unified theory of construct validity. Some theorists disagree with the idea of construct validity as the whole of validity (Kane, 2000; Sireci, 2020), some disagree with the assertion that the consequences of test use should be factored into validity (Lissitz & Samuelsen, 2007), and some have expressed that the complexity of the theory makes it too difficult to use (Royal, 2017; Shepard, 1993).

Kane (2000) proposed an alternative approach to test validity which avoided uses any particular underlying theory for test validity. Kane (2000) stated that the validation of test use must be established through an argumentative process which began with a delineation of "the proposed interpretations and uses of observed test scores" (p. 26) followed by a description of supporting evidence for those interpretations and uses. Mislevy (2009) expanded Messick's inclusion of consequences of test use as necessary for the test validation process by developing a sociocognitive approach to validation which promoted a "constructivist-realist view of validity" (p. 84). In Mislevy's (2009) approach, what is measured is real, but the conception and measurement of what is real is constructed within a particular context and, therefore, may differ widely (Sireci, 2020). In their unified framework of validity and validation, Hubley and Zumbo (2011) reframed Messick's (1989) theory of construct validity by encompassing the theoretical framework within individual and societal values and stressing the need to identify the personal and societal consequences of test use.

Current Approaches to Test Validity

Almost a century after its infancy, experts in the field of test validity still grapple with defining what test validity is and how to establish it (Sireci, 2020). Although individual validity theorists may disagree on nomenclature, theory, frameworks, and approaches, the American Educational Research Association [AERA] et al. (2014) *Standards for Educational and Psychological Testing* (*Standards*) provides the field with a consensus on test validity. Appointed by a joint committee representing the American Psychological Association (APA), the AERA, and the National Council on Measurement in Education (NCME), the authors of the *Standards* have been careful not to commit to any one person's theory, framework, or methodology for test validity. Evident throughout the text are many concepts proposed in Messick's (1989) theory of construct validity, Kane's (2000, 2008, 2013) argumentative approach to validation, and Mislevy's (2009) focus on sociocognitive aspects of validation. According to the *Standards (*AERA et al., 2014), test validity is "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11). Test validation begins with a statement of the test construct—the "concept or characteristic that a test is designed to measure" (AERA et al., 2014, p. 11) and then involves the construction and evaluation of arguments for and against the intended interpretation of test scores and their relevance to their proposed use. Several sources of evidence may support arguments for or against test validity; these sources include evidence about test content, cognitive response processing, internal consistency, relationships with related criteria, and the consequences of testing. Throughout the *Standards 2014*, the authors stress the idea that every step of the assessment process—from test design and development to scoring interpretations and test fairness—ultimately affects the validity of the assessment (AERA et al., 2014).

A Conceptual Framework for Test Validation

According to Sireci (2020), the "AERA et al. (2014) *Standards* provide us with a framework for gathering and organizing validity evidence" (p. 6). Therefore, I adapted the content of the AERA et al. (2014) *Standards* into a conceptual framework for test validation (see Figure 1). To create the framework, I placed at the center the AERA et al. (2014) *Standards'* five "sources of evidence that might be used in evaluating the validity of a proposed interpretation of test scores for a particular use" (p. 13). Surrounding these sources, I placed the additional topics covered in the sections on foundations and operations, which I represented with arrows that encircle and point to the sources of validity. I chose this representation to emphasize that all processes involved in test development, administration, score interpretation, and use of test results impact the validation process. Because AERA et al. (2014) *Standards* chose to include validity, reliability, and test fairness as foundational to test development, interpretation, and use, I shaded those topics in grey. Finally, I outlined in bold dashes the source of evidence that

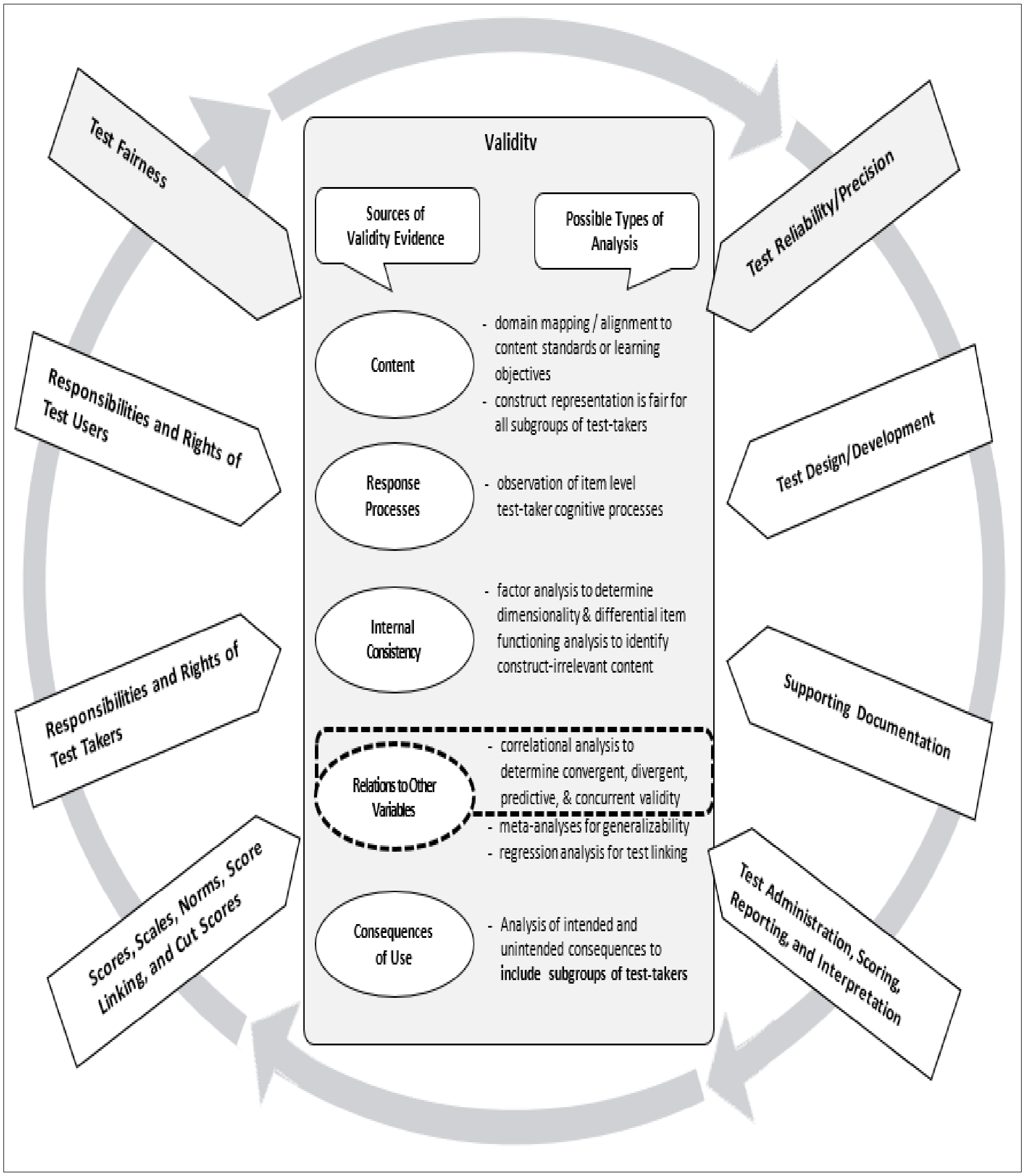

Figure 1. *Standards Test Validation Framework.* This model is adapted from the general structure and content included in the *Standards for Educational and Psychological Testing* by AERA et al., 2014. Chapters 1 through 3 are shaded lightly. Chapter 1, which lists the possible sources of validity evidence, appears in the center of the figure.

directly connected to my research questions, which was correlational analysis used to determine convergent validity under the Relations to Other Variables source of evidence.

Parallels Between Pragmatism and Test Validity

Interestingly, I discovered much overlap in the assumptions of pragmatic philosophy and test validity theory. Pragmatism was born as a philosophy with hopes to end the philosophical debate on the meaning of truth. Peirce (1905), considered to be a pioneer in pragmatic philosophy, expressed his impatience with metaphysical absolutes when he wrote:

> It will serve to show that almost every proposition of ontological metaphysics is either meaningless gibberish—one word being defined by other words, and they by still others, without any real conception ever being reached—or else is downright absurd; so that all such rubbish being swept away, what will remain of philosophy will be a series of problems capable of investigation by the observational methods of the true sciences. (p. 171)

A similar opinion was expressed by Sireci (2020) regarding the contention between theorists who supported Messick's (1989) theory of construct validity and those who did not. In fact, Sireci observed that if:

> We could use many of the tenets of construct validity theory to design and conduct validation efforts, without getting bogged down in the nomenclature . . . the work becomes less philosophical and more practical. Rather than focus on a construct, we focus on testing purposes and specific use of test scores. (p. 9)

A second parallel that exists between pragmatist and modern validity principles includes the pragmatic beliefs on truth being malleable and fallible. For example, in his discussion of modern validity tenets, Royal (2017) asserted that modern validity theorists generally believe that validation is a cumulative, never-ending process to which evidence from multiple investigations is constantly added. This aligns to pragmatic principles of truth as malleable and fallible and Dewey's (1938) assertion that knowledge is a warrantable assertion that only serves as the basis for further inquiry.

A final example of parallelism between pragmatic and perspectives of modern validity theorists is the approach to inquiry that involves the consideration of the context in the development of knowledge. As a specific example, I will compare how both Dewey (1938), using his theory of inquiry, and Cook et al. (2015), using Kane's (2000) argumentative approach for validation, both turned to courtroom scenarios to explain their proposed methods for determining truth. In his *Logic: The Theory of Inquiry,* Dewey (1938) wrote that "the occurrence of a trial at law is equivalent to the occurrence of a problematic situation which requires settlement" (p. 120). Dewey analyzed the thought processes of the judge from the perspective of his theory of inquiry. According to Dewey, the judge first asked himself a question, which was a crucial step, as this question would determine the hypotheses to be considered and what data to accept or reject. The judge then gathered necessary data from witnesses as well as laws relevant to the situation. Finally, the judge tested his hypotheses to render a decision, or in pragmatic terms, a warranted assertion based on his process of inquiry that provided a resolution that benefited all parties to the greatest extent possible.

In an effort to explain their process of test validation, Cook et al. (2015) proposed a courtroom scenario based on Kane's argumentative approach to validation. The authors stressed that the evidence that needed to be collected was dependent on the severity of the crime—i.e., the context—and not only on the facts. The context, in addition to the facts, guided the type of evidence that needed to be gathered and presented to enable the judge to make a decision that is beneficial based on the facts, the context, and the consequences of the decision. Although Cook et al. (2015) viewed the courtroom through the lens of a lawyer rather than a judge, the processes of inquiry proposed by Dewey (1938) and Cook et al. (2015) were similar. In both courtroom scenarios, the evidence collected was dependent upon the question that was asked, which in turn was considered within the context of the crime. The consequences of the courtroom decision were also key in both examples.

Pragmatist theorists Cronbach and Meehle (1955), Kane (2008), Zumbo (2009), and AERA et al. (2014) all proposed pragmatic processes of inquiry in which context and consequences were crucial to the validation process. Zumbo (2009) credited Cronbach and Meehle (1955) with their concept of validity that needed an “explanation rather than the prediction/correlation approach” (p. 74) and asserted that validation was an explanatory process based "on a contextualized and pragmatic view" (p. 73). Kane (2013) justified his argumentative approach to validity, writing that he was not seeking "Truth" (p. 120); rather, in a pragmatic manner, he was seeking “justified belief” (p. 120). Finally, AERA et al. (2014) stated that “effective testing and assessment requirements . . . an awareness of personal and contextual factors that may influence the testing process” (p. 1) that must be considered in the development of a “sound validity argument [that] integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of test scores for specific uses” (p. 21).

The entire focus of my investigation concerns test validity, which is a form of truth. In line with pragmatic principles, modern validity theorists see this truth as contextual and therefore subject to revision. Truth-finding for the pragmatist is never-ending; it is a process of refinement but not of absolute truth. Validity theorists approach test validation in a similar pragmatic lens; validation is a cumulative, never-ending process of refinement that is dependent not only on the test instrument but on the intended as well as unintended consequences of the use of the test.

Chapter Summary

I introduced this chapter with an overview of current issues in adult education student transitions to postsecondary education and the role of college placement testing in postsecondary placement options. I then presented my background of the problem, which was anecdotal information from FF-ABE practitioners that there was a misalignment between TSIA and TABE 11/12 placement levels. Next, I discussed the purpose of this study, which was to compare student performance on the TSIA and TABE 11/12 assessments by investigating relationships between student scores and placements on the two instruments. After I identified my assumptions, the scope of the study, and my research questions with null hypotheses, I defined operational terms pertinent to the investigation's context. I followed reflections on my pragmatic philosophical stance with an overview of test validity theory, which culminated in this inquiry's conceptual framework for test validation. Finally, I discussed parallels between various principles of pragmatism and current views on test validity.

II: REVIEW OF LITERATURE

In this chapter, I present findings from my review of the literature on topics relevant to my investigation. The purpose of my research was to discover relationships between student performance on two tests of academic ability: the Texas Success Initiative assessment (TSIA), which also included an adult basic education (ABE) diagnostic component, and the Tests of Adult Basic Education (TABE 11/12). My interest in comparing students' closely timed performance on the two tests was to determine if there was a positive correlation between the student performance on the two tests. At a practical level, I wanted to see if there was any pattern of overplacement or underplacement between the mathematics TSIA (TSIA-M) ABE Diagnostic and the mathematics Test of Adult Basic Education 11/12 (TABE-M). Understanding consistent relationships between the placements could be helpful for practitioners who use ABE levels from the two assessments to make student placement decisions. I also wanted to know if the relationships between student performance on the two assessments were consistent across various subgroups in the student population, such as gender, race, ethnicity, and age.

My original intent was to analyze relationships on student performance on all areas of the TSIA and TABE 11/12 (reading, writing, and mathematics). However, the number of student records for closely administered TSIA, TSIA ABE Diagnostic, and TABE 11/12 reading and writing tests was insufficient to support inferential statistical analysis. Therefore, I targeted my investigation to explore relationships for the student performance on the mathematics tests only. Because more students enroll in developmental education (DE) mathematics than in any other DE subject, research on the effects of DE mathematics placement on student outcomes is particularly needed (Crisp & Delgado, 2014). As an example, in fall 2019, 53.0% of new community college students scored below the college-ready cut score on the mathematics TSIA, whereas for reading and writing, the percentages who scored below college-ready cut scores were 32.4% and 17.5%, respectively (Texas Higher Education Coordinating Board [THECB], 2020a).

A major tenet of pragmatic thought is the consideration of the impact of one's truths in real-world contexts (i.e., the consequences). The beginning college student experiences the most direct consequences of placement decisions based on standardized test scores. Therefore, I organized the literature view through the experience of the student. Therefore, I first examine the use and validity of high-stakes college admissions and college placement testing. In Texas, advisors place students who score below college readiness on the TSIA in either corequisite DE or DE interventions or refer students to adult education services or the college's career and technical education program. To provide the reader with background knowledge of DE and adult basic education (ABE), I describe the DE and ABE historical context and current landscape nationwide and in Texas. Finally, because my investigation revolves around the TSIA's use for placement, I include a section on assessment and placement practices for DE and ABE.

Use and Validity of College Admissions and Placement Testing

According to the American Educational Research Association's (AERA) position statement on high-stakes testing, it is not the test itself but the uses of the test results that create a high-stakes test. If tests "carry serious consequences for students or for educators," they are high-stakes tests (AERA, 2000, para. 3). Specifically, AERA cautions that the use of a test with poor reliability and validity for high-stakes outcomes can bring the “potential for serious harm” (2000, para. 4). The SAT and ACT college admissions tests are considered high-stakes tests as scores on these tests are weighed heavily by the majority of 4-year institutions of higher education (IHEs) for admissions (What Works Clearinghouse, 2016).

Just like SAT and ACT college admissions tests, college placement tests are high-stakes tests with high-stakes consequences (Castro, 2013; Scott-Clayton, 2012). According to many researchers, the consequences of scores on college placement tests are quite significant in terms of the additional cost and time that students will incur if their scores result in DE placement (Bailey et al., 2010; Horn & Nevill, 2006; Melguizo et al., 2008). Based on their test scores from college placement tests, the majority of college freshmen must enroll in developmental coursework. National estimates for students who take at least one DE course during college range from 40% to 60% (Bailey et al., 2010; Chen & Simone, 2016). For public 2-year institutions, this estimate is over 65% (NCES, n.d.).

In 2011, 100% of the 970 public 2-year institutions in a national survey were using a mathematics test score, and 94% were using a reading test score for placement purposes (Fields & Parsad, 2012). Proponents of using an assessment with standard cut-off scores for placement have suggested that this assists in the alignment of secondary to postsecondary college readiness standards, facilitates student transition to postsecondary, and helps maintain rigorous academic standards (Melguizo et al., 2016; National Center for Public Policy and Higher Education & Southern Regional Board, 2010). However, the publisher of the most widely used college placement test, the College Board's

ACCUPLACER (Barnett & Reddy, 2017), has advised colleges to use additional measures along with test scores to evaluate and place incoming students (College Board, 2018a). It appears that many IHEs have responded to the call for additional measures. As of 2018, 57% of 1,000 non-selective and 2-year IHEs surveyed reported using other factors, such as high school grade point average (HSGPA), in conjunction with college placement test scores for mathematics placement; 51% did the same for reading and writing placements (Rutschow & Mayer, 2018).

SAT and ACT Validity Research

I have included validity research on the SAT and ACT, which are college admissions tests rather than college placement tests, as a baseline comparison to research on college placement test validation studies. Typically, the validity of a college admission test such as the SAT and ACT depends on its ability of test scores to predict students' first-year college GPA (Zwick & Sklar, 2005). Several investigators have reported support for the predictive validity of SAT and ACT scores by correlating test scores with first-year college GPA, especially when test scores are combined with HSGPA (Bettinger et al., 2011; Sackett et al., 2009; Westrick et al., 2019). The College Board's most recent predictive validity for the SAT reported an observed correlation of .32 between combined SAT scores and college first-year college GPA. After statistical corrections were introduced to correct issues in sample size, range restriction, and criterion reliability, researchers reported a correlation of .51, considered to be strong in terms of predictive validity (Westrick et al., 2019). ACT's most recent predictive validity study reported an observed correlation of .38 between combined ACT scores and first-year college GPA.

After corrections were made for sample size and range restriction, the correlation increased to .51 (Westrick et al., 2015).

However, a recent investigation of the SAT suggested that predictive validity varied across subpopulations in individual 4-year IHEs, thereby creating subgroups of test-takers who are advantaged or disadvantaged by their SAT scores when scores are used for admissions decisions. Furthermore, differences in high school and college grading practices and the composition of coursework at the local level may unfairly influence predictive validity (Aguinis et al., 2016). The College Board confirmed this subpopulation variance with its own research, stating that SAT scores used alone may overpredict first-year college GPA for African American and Hispanic students and underpredict first-year college GPA for females. The addition of HSGPA to SAT scores minimizes these effects (Mattern & Patterson, 2013). Noble and Sawyer (2013) also observed that when used as a single measure, HSGPA served as a better predictor of student performance than the ACT used alone. The correlation of HSGPA to first-year college GPA was .48, whereas the correlation of composite ACT scores to first-year college GPA was .44. However, combining the composite ACT score with HSGPA increased predictive ability for a first-year college GPA of a C or above to .54.

Due to the variability of predictive validity at the local institution level and subpopulation variance, both the College Board and ACT endorse using student test scores along with other measures, such as HSGPA, for admissions decisions (ACT, 2020; College Board, 2021). However, only about one half of IHEs appear to be performing local predictive validity studies; a 2015 survey of over 1,600 IHEs indicated that though 78% required SAT or ACT scores for admissions, only 51% conducted local SAT or

ACT predictive validity studies (National Association for College Admission Counseling, 2016).

College Placement Test Validity

I found the literature to be mixed in researchers' conclusions regarding the predictive validity of college placement tests. Overall, I found that researchers associated with test publishers tended to report more positive results regarding the predictive validity of their testing instruments than did third-party researchers. However, I also noted that publisher-sponsored validity research used nationwide data for their population samples, whereas independent researchers exploring predictive validity used data from either a single IHE or a statewide IHE system. According to the AERA et al. (2014) *Standards in Educational and Psychological Testing:*

> Validation is the joint responsibility of the test developer and the test user. The test developer is responsible for furnishing relevant evidence and a rationale in support of any test score interpretations for specified uses intended by the developer. The test user is ultimately responsible for evaluating the evidence in the particular setting in which the test is to be used. (p. 13)

Therefore, because the test publishers demonstrate predictive validity for national rather than local or statewide IHEs, it is expected that the predictive validity and placement accuracy will differ based on the local population sampled (Mattern & Packman, 2009). For this reason, the College Board offers the "Admitted Class Evaluation Service (ACES) . . . a free online service for higher education institutions that predicts how ad mitted students will perform at your institution generally and how successful they can be in specific courses" (College Board, 2021, p. 3).

The TABE 11/12 was not designed for college placement and is therefore not included in my discussion of college placement tests. Federally-funded adult basic education (FF-ABE) programs require the TABE as a pre-and post-test. FF-ABE services are always free for instruction under high school academic levels and free or very low-cost for instruction at the high school level. Because classes are not credit-based, students may move up or down levels as agreed upon by program staff and the student. Therefore, the TABE 11/12 as used in the FF-ABE system is not a high-stakes assessment.

Currently, at least two states have accepted TABE 11/12 scores as alternative evidence of college readiness, and the publishers have been asked by several states to assist with concordance tables (Ellis, n.d.; Johnson, 2019.; New Mexico Public Education Department, 2017). The conditions which would make the TABE 11/12 test a high-stakes test similar to a college placement test would most likely exist only in FF-ABE programming that transitions students to college at college readiness levels through TABE 11/12 scores. However, in the context of this study in Texas, the TABE 11/12 not used as a high-stakes assessment.

Predictive Validity Research

Researchers choose various types of analyses to establish predictive validity. First, researchers must select the criterion against which the test score is evaluated. Typically, researchers select the student's success in their first English, math, or reading-intensive course, with indicators of course success set to a particular grade earned in the course. Researchers may also choose use regression analyses that determine the percentage of the course grade that can be explained by the test score. For placement tests, most researchers provide the accuracy of the student placement into the college-level courses, with correct placement defined as the percentage of students who were placed in the course who passed the course based on the criterion of course success (Mattern & Packman, 2009; Scott-Clayton et al., 2014).

In addition to the percentage of correct placements, researchers may choose to provide a breakdown of the percentages of students who were overplaced, which occurs when students fail to meet the college-level grade criterion, or underplaced, which occurs when students who were placed into a DE class were predicted to have passed the associated college-level course based on their test scores. To determine overplacement and underplacement, researchers need to compare a group of students who were placed by the same test scores into DE or into college-level coursework. This happens naturally when institutions use different sets of cut scores to determine college readiness levels; i.e., a score of 385 may determine math college readiness at one set of institutions and a score of 380 may determine math college readiness at another set of institutions (Scott-Clayton et al., 2014).

Having students with the same test scores in both DE and college-level courses also occurs when institutions recommend placement rather than require placement based on test scores, which results in student non-compliance. In this scenario, some students who were recommended to enroll in DE chose to enroll in college-level courses, and some students who were recommended to enroll in college-level coursework chose to enroll in DE (Mattern & Packman, 2009; Scott-Clayton et al., 2014). Finally, as in cases similar to Texas, students may be exempted from placement into DE despite having test scores that indicate DE placement. In Texas, there are several exemptions which qualify students as exempt from meeting minimum TSIA scores, such as having met minimum scores on other designated standardized exams, passing a high school college preparation course, prior success in postsecondary, or veteran status (19 Tex. Admin. Code §4.54).

According to Mattern and Packman (2009), the "main goal of placement testing is to enroll students in courses that are aptly challenging to their current knowledge level so as not to bore or frustrate, which can lower motivation to perform" (p. 1). When conducting a predictive validity investigation, rather than validating the assessment itself, researchers are actually examining the "validity of the cut scores, or the test scores required for placement into one course … over another course" (p. 2). Cut scores, which are the values on the scaled scores that classify students into two or more proficiency levels, are usually initially determined through a standard-setting process. The validation argument for cut scores derived from a standard-setting procedure "tends to come from three places: procedural evidence, internal consistency evidence, and external evidence" (McClarty et al., 2013, pp. 80–81).

The standards-setting process typically begins with test items that have been field-tested along with proficiency level indicators based on curricula standards. To the extent possible, external empirical data should be collected and distributed to several pre-selected content experts prior to the standards-setting process. This data might include concurrent and predictive validity research along with impact data, which is the percentages of test takers who would be classified in each level as based on pilot testing given particular sets of cut scores (DRC, 2017; McClarty et al., 2013). Next, subject matter experts go through two or more rounds of standards-setting in which they identify the proficiency levels for the test items using one of several cut-score methods, such as the commonly used bookmark method. After empirical data has been collected, the cut scores should be regularly validated and revised as necessary to ensure that the desired percentages of students are placing into the test's proficiency levels (McClarty et al., 2017; Morgan & Michaelides, 2005).

Publisher-Sponsored Test Validation Research

Until the 2015 retirement of the COMPASS (Fain, 2015), the College Board's ACCUPLACER and ACT's COMPASS were the two most commonly used college placement tests through the nation (Rutschow & Mayer, 2018). The College Board is also the developer of the Texas TSIA, which is a customized ACCUPLACER (College Board, 2014). Mattern and Packman (2009) and Cui and Bay (2017) conducted College Board-sponsored investigations of the placement validity of the ACCUPLACER and the TSIA, respectively. On behalf of its publisher, ACT, Westrick and Allen (2014) examined the predictive ability of the COMPASS.

Mattern and Packman (2009) used a meta-analysis design to evaluate the ability of the ACCUPLACER to predict students' grades of a B or above and of a C or above in their first English, math, and reading-intensive courses. Statistical analysis included the calculation of observed correlation coefficients, operationalized correlation coefficients that were corrected due to restricted range and criterion reliability, and placement accuracy rates. Based on the average of a 70% placement accuracy across all ACCUPLACER tests, Mattern and Packman (2009) concluded that the ACCUPLACER exhibited sufficient predictive validity for placement purposes.

In contrast to Mattern and Packman's (2009) approach, Westrick and Allen (2014) explored the ability of COMPASS scores to predict students' grades of a B or above using hierarchical logistic regression. Westrick and Allen (2014) also reported the placement accuracy rates for each COMPASS test. The researchers compared the predictive validity of the COMPASS to the predictive validity of using high school grade point average (HSGPA) alone or in tandem with COMPASS test scores. The authors concluded that HSGPA generally a better predictor than COMPASS test scores and that in all cases, using COMPASS test scores with HSGPA increased predictive ability.

In 2017, THECB contracted the College Board, the test developer and test publisher of the TSIA, to conduct a predictive validity student for the TSIA. the College Board (Cui & Bay, 2017). As did Mattern and Packman (2009) with their ACCUPALCER research, Cui and Bay (2017) reported correlations between the tests and the associated college-level gateway courses. In contrast to Mattern and Packman (2009), the authors provided underplacement and overplacement rates in addition to placement accuracy rates. Finally, by conducting logistic regression analyses, Cui and Bay (2017) were able to report the probability that a student with a minimum college readiness cut score would actually pass the college-level class with a C– or above.

Due to the differences in research design and analyses, it is somewhat difficult to directly compare the predictive ability results from test publisher-sponsored research. However, in Table 1, I list each publisher-sponsored study, the test or combination of test scores analyzed as the predictor variable, the college-level course grade or grades criterion, the number of student test scores examined, and the type of analysis used by the researchers. The most notable difference is the ACCUPLACER's much higher 77.3% correct placement rate for the college-level course criterion of a C or above reported by Mattern and Packman (2009) compared to the TSIA's 62.55% rate. Excluding the 77.3%

Table 1

Publisher-Sponsored Research for the ACCUPLACER, COMPASS, and TSIA

Research Study	Test or Test Combination	College-Level Course Criterion	*n*	Type of Analysis Correlation (*r*)	Logistic Regression Probability (*P*)	Correct Placement Rate (%)
Mattern and Packman (2009)	ACCUPLACER Average of 3 Math Tests	B or above C or above	11,266 11,266	.31 .27	– –	65.8 77.3
Westrick and Allen (2014)	COMPASS Average of 5 Math Tests	B or above	18,927	–	.51 [a]	68.8
Cui and Bay (2017)	TSIA Math	B– or above C– or above	3,690 3,690	.26 .21	.64 .40	– 62.6

Note. Unless otherwise noted, *r* and R^2 values were reported as significant by the researchers. [a] The logistic regression probability (*P*) was calculated from the researchers' reported unstandardized fixed-effects *B* using the following formulas: $OR = (\exp)^B$ and $P = (OR) / (1 + OR)$.

outlier, however, the rate for students to pass the college-level course with a C– or above ranges from 62.6% to 66.6%.

Another item to note is that the magnitude of the correlation or logistic regression probability does not have a clear relationship with the correct placement rates. Correlation establishes the existence of a non-causal relationship between student scores and course grades and regression establishes the existence of a causal relationship between student scores and course grades; both correlation and regression results help validate the predictive ability of the assessments themselves (AERA et al., 2014;

Kosiewicz & Ngo, 2020). However, correct placement rates rely on the cut scores used for college readiness and validate the cut scores used for placement rather than the assessment. According to the *Standards* test validation framework (see Chapter 1), evaluation of the correct placement rates for the assessments falls under the Consequences of Test Use source of validity.

College Board-sponsored validation for the ACCUPLACER and the TSIA (Cui & Bay, 2017; Mattern & Packman, 2009) did not provide comparisons of the predictive ability of test scores as opposed to HSGPA. There was also no attempt to analyze group differences, which would have offered a source of validity evidence under Consequences of Use in the *Standards* test validation framework (see Chapter 1). In contrast, the research design for ACT's validation research (Westrick & Allen, 2014) broadened the context for validation by including comparisons of the predictive validity of COMPASS test scores alone, HSGPA alone, and in combination with HSGPA. The researchers also chose to analyze differences in the COMPASS's role in predictive validity for both traditional students (19 years and under) and nontraditional students (20 years and older). In doing so, Westrick and Allen (2014) were able to discover that accuracy of placement rates for nontraditional students, whether determined by using COMPASS scores alone, HSGPA alone, or combined COMPASS scores and HSGPA, was lower for all tests than the placement accuracy rates of traditional students. This finding would be important in test score interpretation for students who are older. Due to the evidence that HSGPA was generally a better predictor than the COMPASS tests for college gateway course success, ACT announced that they would discontinue the test at the end of 2016, writing that "analysis of customer feedback, empirical evidence, and postsecondary trends led us to conclude that ACT COMPASS is not contributing as effectively to student placement and success as it had in the past" (Fain, 2015, para. 8).

Independent Validation Research

In contrast to research on the ACCUPLACER, COMPASS, and TSIA assessments conducted by their publishers, several researchers have concluded that these tests had limited predictive validity on future student academic performance such as grades in DE courses and college gateway courses, college GPA, credits earned, retention, and persistence (Belfield & Crosta, 2012; Medhanie et al., 2012; Melguizo et al., 2016; Ngo & Melguizo, 2016; Scott-Clayton, 2012; Scott-Clayton et al., 2014). Belfield and Crosta (2012), Scott-Clayton, (2012), and Scott-Clayton et al. (2014) focused on identifying and analyzing the rates of what they termed severe underplacement and overplacement. Scott-Clayton et al. (2014), Ngo and Melguizo (2016), Melguizo et al. (2016), and Barnett et al. (2020) compared placement rates across demographic groups. Finally, several researchers compared placement rates from using test scores only HSGPA only, and both measures together (Bahr et al., 2019; Belfield & Crosta, 2012; Scott-Clayton et al., 2014).

Hughes and Scott-Clayton (2011) explored issues with college placement, especially for open access community colleges that were typically using a single college placement test score to assign students to college-level coursework or to various levels of DE. The authors concluded that the widely used COMPASS and ACCUPLACER tests were "reasonably valid predictors of students’ grades in college-level coursework, but the placement recommendations that result from the use of these tests do not clearly improve student outcomes" (pp. 343–344). The authors recommended that research be performed that examined alternative types of placement policies which would include additional cognitive as well as non-cognitive measures.

Placement Using Alternate Assessments. Medhanie et al. (2012) compared the predictive validity of the ACT versus the ACCUPLACER for students in a community college system. The authors used logistic regression to determine the percentage of variance in DE and subsequent college gateway courses explained by the ACT and the ACCUPLACER. The authors found that the ACCUPLACER provided little or no value over the ACT toward the prediction of grades in the student's DE and subsequent gateway math courses. Also looking at the ACCUPLACER, Ngo and Melguizo (2016) examined differences in the predictive ability of a diagnostic math test, the Mathematics Diagnostic Test Project (MDTP), and the online ACCUPLACER math test. When colleges changed their math placement test from the MDTP to the ACCUPLACER, placement error rates increased, especially for students who had scored in a band just below the college readiness cut score for math. The authors concluded that that "diagnostics may be better able to identify student skill than computer-adaptive tests" (p. 23).

Placement Using Multiple Measures. Several researchers have conducted research that compared how well placement policies that used test scores alone, HSGPA alone, or a combination of the two measures predicted various student outcomes such as DE and college-level gateway course grades, college GPA, credits earned, persistence, or completion. In doing so, many researchers have reported discovered that the use of HSGPA alone resulted in more accurate placement rates for the majority of students than the use of test scores alone. However, researchers determined that combining test scores with HSGPA provided the most accurate placements (Barnett et al., 2020; Belfield & Crosta, 2012; Ngo & Kwon, 2015; Scott-Clayton, 2012; Scott-Clayton et al., 2014). In turn, increased placement rates were "associated with lower rates of misplacement and higher rates of enrolling in and succeeding in college-level courses in math and English" (Belfield & Crosta, 2012, as cited in Scott-Clayton et al., 2014, p. ES-1).

Placement Across Demographic Groups. Scott-Clayton et al. (2014) examined the effects of assessment and placement policies on student placement based on demographic characteristics. The authors reported that using HS transcripts alone for math placement would have increased the representation of female and Hispanic students while decreasing the representation of Black and Asian students in college-level math gateway courses. For English placement, using HS transcripts alone would have almost tripled the representation in college-level English classes for Asian students while reducing the percentage of Black students in those classes by about 50%. More recently, Barnett et al. (2020) conducted a large-scale study that randomized student placement based on ACCUPLACER test scores only with student placement based on ACCUPLACER scores, HSGPA, and years since high school graduation.

Tracking students over three semesters, the authors reported that except for males in math, all other student populations saw an increase in enrollment into college-level courses. Furthermore, the authors documented a significant increase in gateway math student success when using the multiple measures placement for students who were female, non-Pell recipients, or White. However, they reported a negative effect from using the multiple measures placement on Black students' gateway math course success and a null effect for students who were Pell recipients or Hispanic. In contrast to math, using the multiple measures placement for English gateway courses resulted in a significant increase in student success across all student demographic groups.

TSIA Validity Research. A search of the literature revealed no peer-reviewed literature that examined TSIA validity. However, I located three dissertations that analyzed the effects of the TSIA on student math-related outcomes, including DE placement, retention, and gateway math course grades. I also found two dissertations that dealt with the faculty's perception of the TSIA as a placement instrument and one study that examined characteristics of students who tested as below college-ready on the TSIA. Below I summarize these TSIA-related dissertations.

Three dissertations included an analysis of TSIA math scores to gateway math course grades. Carter's (2019) study consisted only of university students who had enrolled directly into gateway math courses (and were therefore considered college-ready). Carter found a significant uncorrected correlation of .25 between TSIA math scores and gateway course grades. After controlling for age, ethnicity, gender, family income, and parent education level, the TSIA math scores had a significant positive effect on gateway math course grades. Lee (2018) investigated the predictive validity of TSIA mathematics scores on community college STEM majors' college algebra course grades and STEM retention. After controlling for age, gender, and HSGPA, the TSIA-M scores explained 13% of the variance in college algebra course grades and 7% of the variance in STEM mathematics retention. Maxwell (2021) examined the ability of the TSIA mathematics score alone, cumulative HSGPA alone, or combined TSIA mathematics score and HSGPA to predict community college students' first gateway mathematics course grades in a community college setting during fall 2019. Regression analyses yielded HSGPA as the only significant predictor of course grades, with HSGPA contributing to the variance in test scores by 8.1%.

Maxwell (2021), Carter (2019), Shields (2016), and McAdams (2017) all recommended incorporating noncognitive factors such as HSGPA or HS transcripts for placement purposes. Evidence from research on the ACCUPLACER, COMPASS, and TSIA strongly suggests that HSGPA, even if it is self-reported, should be considered as a second measure to increase placement accuracy (Carter, 2019; Maxwell, 2021; Ngo & Melguizo, 2016; Scott-Clayton et al., 2014; Westrick & Allen, 2014). In the *Standards* test validation framework, a source of validity evidence comes from the consequential use of the test. Using the test alone without HSGPA does not infer that a test score is not valid, but it does provide evidence that the failure to use HSGPA with the test score is, in fact, a misuse on the part of the test users.

In addition to quantitative research that provided evidence for the TSIA, I located two dissertations that investigated faculty perceptions of the TSIA as a measurement of college readiness. In both studies, the authors used structured interviews with DE faculty and analyzed their responses for emerging themes. Shields (2016) noted an emerging theme that the TSIA was an improvement over prior exams but still inaccurately placed students within DE levels. Another theme included the concern that some students may not have taken the exam seriously. All ten professors interviewed believed that non-cognitive factors need to be considered with placement decisions.

Similar to Shields (2016), McAdams (2017) interviewed DE faculty to assess their perceptions of the TSIA as a placement tool. McAdams reported that several professors believed that the mathematics TSIA cut scores used for placement into various DE levels were not well aligned with the DE mathematics curriculum. As with findings from Shields (2016), faculty overwhelmingly indicated a need for non-cognitive factors to be considered for placement purposes. The acceptance of the validity of an exam for placement purposes is part of the validation process. Faculty perceptions of placement were critical, with most professors believing that students had been placed higher than their abilities and that cut scores needed to be raised. However, results from quantitative studies regarding placement have suggested students are underplaced into DE and many who were placed in DE could have passed college-level coursework. Therefore, instead of raising cut scores, several researchers have suggested that cut scores should be lowered or adjusted based on thorough review and analysis of institutional data in order to raise academic success for all students (Barnett et al., 2020; Melguizo et al., 2016; Ngo & Melguizo, 2016).

Stout (2014) used data from over 6,000 students to analyze factors that increased the probability that a student taking the TSIA would not test as college ready. The author found that Black students were 2.5 times more likely to test below college readiness levels than White students. The author also reported that as time since prior school increased, the probability of testing below college ready increased and that as parent educational levels increased, the likelihood to test as below college decreased. The results of this study align with Daugherty et al.'s (2019b) study on characteristics of students who were placed at ABE levels 1 through 4 on the TSIA. The results from both Daugherty et al.'s (2019b) and Stout's (2014) studies provided evidence for the Consequences of Use category of the *Standards* test validation framework by examining how the use of the TSIA may be creating a “vacuum in which special populations of students are being lost, disenfranchised, or willingly excluded from educational opportunity” (Stout, 2014, p. 95).

Findings from research on the TSIA were predominantly more negative than positive. Carter (2019) and Lee (2018) reported statistically significant contributions of the TSIA scores to students’ college-level algebra courses. However, Lee (2018) reported that after controlling, gender, ethnicity, and age, and HSGPA, the effect of the TSIA-M score on course success size was only 7%. Because Carter’s (2019) used a population of 4-year university students, they may have had higher math skills overall than the student samples used in Lee’s (2018) and Maxwell’s (2021) studies. The only positive theme that emerged from the two studies on faculty perceptions of the TSIA was that the test was an improvement over prior assessments.

Developmental Education

Shifting away from my review of college placement testing and validity, I now discuss literature related to developmental education (DE). Postsecondary institutions of higher education (IHEs) typically require incoming students with no evidence of college-level academic skills to enroll in prerequisite or corequisite DE coursework. Because community colleges are primarily open access and typically accept all students regardless of their basic skills proficiency, most students who enroll in DE attend public 2-year colleges (Atkinson & Geiser, 2009; Desai, 2012). Although developmental education (DE) may be broadly defined as any form of student learning assistance support (Boylan & Bonham, 2007), most postsecondary stakeholders view DE as basic skills instruction provided to or required for students deemed as below-college ready by their IHEs.

According to Arendale (2011), defined broadly, developmental education (DE) involves academic support for students, and with this definition in mind, DE includes individual and group tutoring, computer-assisted self-directed learning, and classroom instruction designed to improve academic performance. Throughout related scholarly literature, many researchers refer to DE as remediation, as do many members of the public, IHE faculty and staff, and even DE practitioners (Arendale, 2005; Holschuh & Paulson, 2013; Stahl & Armstrong, 2018). However, pioneers in the field, such as Arendale (2011) and Boylan et al. (2017), contrast remediation and DE in several ways. First, remedial instruction uses a deficit model in which the student has skills gaps that need to be diagnosed and filled, whereas DE attempts to provide a learning environment that allows students to develop skills that they need for collegiate success (Arendale, 2011). Stand-alone classes that teach basic skills without providing supportive student services are often remedial in nature. In contrast, a DE class offers basic skills within programming that provides an "integration of courses and services governed by the principles of adult learning and development" (Boylan et al., 2017, p. 2).

Around the turn of the twenty-first century, a push for higher college completion rates led to scrutiny of the efficacy of developmental education (Attewell et al., 2006; Schak et al., 2017) as well as the implementation of a wide variety of reform initiatives concentrated on one or more of the following areas: accelerated instruction, and instructional strategies, and assessment and placement. Accelerated approaches include course redesigns that allow students to enroll in and pass college-level coursework more quickly by compressing content, requiring corequisite academic support concurrent to college-level coursework, and designing math pathways that provide math coursework specific to a student's degree plan (Daugherty et al., 2021). Instructional strategy reforms, often paired with course redesign, strive to engage students and may include contextualized or technology-enhanced courses (Rutschow et al., 2019). Finally, assessment and placement reforms include diagnostic and curriculum-aligned assessments, using more than one assessment for placement decisions, and changes to advising (Hodges et al., 2020; Ngo & Melguizo, 2016).

Developmental Education: Early History

For the first two centuries of higher education in the United States, academic support occurred in the form of tutoring. However, the after mathematics of the Civil War created demand on the part of Blacks and women to achieve postsecondary education, and federal legislation encouraged the establishment of IHEs. To address the many students who needed academic assistance, many IHEs housed separate academic preparatory programs. In the mid-1900s, federal legislation in response to World War II and the civil rights movement created increasing opportunities for many more Americans to enroll in colleges. Although the open doors admissions policies, especially at junior or community college, made it possible for almost any adult to attend higher education, these policies also resulted in large numbers of students who needed academic support. Therefore, many IHE had "to introduce remedial courses into the college curriculum to attempt a redress of their inadequate academic skills and knowledge" (Arendale, 2011, p. 71).

In the 1970s, two new academic support models were introduced: the learning assistance center and Supplemental Instruction. These two models both avoided using the term remedial and ensured that academic support was available for all levels of students rather than just for those who lacked college-level skills. Educators in the field of developmental education also moved to more broadly define and rename their work. An early definition describes DE as a program "to develop the diverse talents of students . . . to give attention to the fullest possible development of talent and to develop strengths as well as to correct weaknesses" (Cross, 1976, as cited in Arendale, 2005, p. 72) developed into a 1996 definition that is still quite accurate today:

> A field of practice and research with a theoretical foundation in developmental psychology and learning theory. It promotes the cognitive and affective growth of all learners, at all levels of the learning continuum. It is sensitive and responsive to the individual differences and special needs among learners. (National Association of Developmental Education, 1996, as cited in Arendale, 2005, p. 73)

Developmental Education: National Prevalence

Incoming college students of all genders, ages, races, ethnicities, and socioeconomic levels enroll in DE. However, Students referred to DE differ from those placed in college coursework in "gender, ethnicity, first-generation status, academic preparation and experiences during high school, and delayed college entry" (Crisp & Delgado, 2014, p. 99). Depending on the state or the individual institution, student participation in DE participation may be required or optional. DE services may also appear as stand-alone basic skills classes or short-term instructional interventions that students must complete prior to college-level coursework or as support courses, peer tutoring, or other types of instructional interventions offered concurrently with college-level coursework, such as the corequisite model (Whinnery & Pompelia, 2019).

National estimates for college DE enrollment vary depending on survey methodology and institutional classification, with approximately 40%–50% percent of 4-year IHE students and 55%–68% who enroll in 2-year IHE participating in DE at some point in their college careers (Attewell et al., 2006; Bailey, et al., 2010; Chen & Simone, 2016). Between 2003 and 2009, approximately 68% of beginning college students at public 2-year institutions and 40% of beginning college students at public 4-year institutions enrolled in one or more DE courses. Of these students, almost 60% at 2-year institutions enrolled in DE math, while about 28% enrolled in DE English. Four year-colleges experienced an enrollment rate of approximately 33% in DE math and 11% in DE English (Chen & Simone, 2016).

Using recently released data from the Beginning Postsecondary Longitudinal Study 12/17 (BPS:12/17), it was possible to obtain estimates of DE enrollment based on student transcripts rather than surveys for first-time-in-college (FTIC) students who enrolled in fall 2012 (NCES, n.d.). Table 2 shows a comparison of results provided by Chen and Simone's (2016) analysis of NCES BPS:03/09 DE enrollment data with NCES BPS:12/17 DE enrollment data. DE enrollment rates from the two sets of data are very similar for students who enrolled in DE at public 2-yearr institutions, but at 4-year public IHEs, total student enrollment in DE courses dropped almost 10% (Chen & Simone, 2016; NCES, n.d.).

A second source of national DE participation data is available from Complete College America (CCA). The CCA data dashboard provides statistics on DE participation at national and state levels disaggregated by age, gender, race/ethnicity, and Pell grant status (CCA, n.d.) Table 3 displays NCES BPS:12/17 data for first-time-in-college

Table 2

National NCES 2011–2012 and CCA 2016 DE Math Enrollment by Subject Area

Subject Area	2-Yr Public IHEs (%)		4-Yr Public IHEs (%)	
	NCES [a] BPS:03/09	NCES [b] BPS: 12/17	NCES [a] BPS:03/09	NCES [b] BPS: 12/17
DE enrollment				
In any subject area	68.0	65.2	39.6	29.4
In math	59.3	54.6	32.6	24.4
In English	28.1	28.0	10.8	8.5

Note. [a] Chen and Simone (2016) reported the percentage of students surveyed who indicated they had ever enrolled in a DE course in any subject area, a DE math course, or a DE English course between 2003 and 2009. [b] NCES BPS:12/17 data was obtained through the NCES PowerStats (NCES, n.d.).

students in fall 2011 and the same CCA data from fall 2016. Although it was not possible to directly compare the data from the two sources due to differences in the data collection methods and years (2011 and 2016), I analyzed the data to look for possible trends in the DE participation rates for distinct student populations. My examination revealed that DE decreased more rapidly than other student groups. In 2-year colleges, the rates of DE math enrollment over age categories remained relatively level, whereas in 4-year universities, the rates of DE math enrollment more than doubled as age increased.

Developmental Education: Prevalence in Texas

Data on DE participation is available through annual DE reports for each of the three higher education systems: 2-year community colleges, 2-year technical colleges, and 4-year or more universities. Data for fall cohorts of first-time-in-college (FTIC)

Table 3

National NCES 2011–2012 and CCA 2016–2017 DE Math Enrollment by Student Characteristics

	2-Yr Public IHEs (%)		4-Yr Public IHEs (%)	
Student Characteristic	NCES [a] 2011–2012	CCA [b] 2016	NCES [a] 2011–2012	CCA [b] 2016
Total DE math enrollment	55.0	52.0	27.3	24.0
By gender				
Female	56.1	55.0	30.2	24.0
Male	53.6	47.0	23.9	21.0
By race/ethnicity				
Asian	49.6	33.0	18.7	14.0
Black or African American	64.4	61.0	45.1	37.0
Hispanic	55.0	50.0	34.1	27.0
White	52.9	44.0	23.1	19.0
Age				
19 and below	53.7	51.0	24.6	22.0
20 to 24	59.1	47.0	47.4	39.0
25 and above	57.3	53.0	66.2	53.0

Note. [a] The NCES 2011–2012 data drew from a sample of 15,300 students and is based on student transcripts. NCES data was sourced from BPS:12/17 (NCES, n.d.). [b] CCA data was sourced from the Complete College American Data Dashboard; the sample size is unknown (CCA, n.d.). A metrics technical guide for the data is available (CCA, 2017).

students include initial college readiness status in mathematics, reading, and writing and enrollment. DE enrollment rates at Texas community colleges were much greater than those at 4-year colleges. In math DE, enrollment was approximately four times greater, and for reading and writing, about six times greater. Students at both 2- and 4-year IHEs had higher success rates for DE reading than for DE math or DE writing, which had with outcome measures of first-and second-year gateway success rates. The report also provides DE enrollment per subject area with related DE course pass rates as outcome measures for year 1 and year 2. Finally, two-year student persistence rates are available in the report (THECB, n.d.-a).

Table 4 provides data available for 2017 first-time-in-college (FTIC) students who enrolled below college readiness per subject area, as well as data on DE enrollment and course success rates. DE enrollment appears both as a percentage of the total student enrollment and as a percentage of students who entered below college readiness in the subject area. At both IHE levels, students frequently delayed taking DE courses, more so in reading and writing than in math. Also, at both institutional levels, enrollment in DE passing rates that were similar within each type of IHE. Finally, students at 4-year colleges had higher success rates than those at 2-year colleges, particularly in DE math.

Compared to national DE data, DE enrollment rates of 2018 FTIC students in Texas were lower than those reported nationally in 2011 and 2016 for all public institutions (see Tables 3 and 4). However, at both the national and Texas level, regardless of institution type, DE math enrollment exceeded enrollments in DE English, reading, or writing. Furthermore, there were much greater gaps in DE enrollment rates between Texas 2- and 4-year colleges than differences between 2- and 4-year college DE enrollments reported in the national data. For example, DE math in Texas community colleges had an enrollment rate four times greater than DE math enrollment rate at 4-year colleges. In contrast, national DE math enrollment at 2-year colleges was only two times greater than at 4-year IHEs.

Table 4

College Readiness Status and DE Enrollment for First Time in College Student Fall 2018 Cohorts in Texas by Subject Area and Type of Institution

Student Characteristic	2-Year Public IHE [a]		4-Year Public IHE	
	N	%	*N*	%
Total statewide enrollment	109,143	100%	90,144	100%
Entered below college ready				
In all areas	14,497	13.3%	497	.5%
In one or more areas	63,034	58.8%	12,772	14.2%
In math	57,356	52.6%	11,647	12.9%
In reading	34,677	31.8%	4,613	5.1%
In writing	19,074	17.5%	2,353	2.6%
Percentage of all cohort students enrolled in DE courses in Year 1				
Math DE	42,472	38.9%	9,573	10.6%
Reading DE	22,801	20.9%	2,821	3.1%
Writing DE	9,613	8.8%	1,047	1.2%
DE enrollment and success				
Enrolled in DE math	42,472	74.0%	9,573	82.3%
Passed DE math	18,506	43.6%	5,867	61.3%
Enrolled in DE reading	22,801	65.8%	2,821	61.1%
Passed DE reading	14,024	61.5%	2,097	74.3%
Enrolled in DE writing	9,613	50.4%	1,047	44.4%
Passed DE writing	5,115	53.2%	662	63.2%

Note. Data was sourced from THECB Developmental Education (TSI) annual reports (THECB, 2018a, 2018b). [a] Data does not include two 2-year technical colleges with FTIC fall 2018 enrollments totaling 3,870 students.

THECB also provides a 3-year longitudinal report for community colleges and a 6-year longitudinal report for universities that compares persistence and graduation rates of students who needed DE and those who did not. At 2-year public colleges, 3-year graduation rates have been increasing steadily for both students who needed DE and those who did not since the TSIA testing began in 2013. However, at 4-year institutions, graduation rates for both groups of students have remained fairly level. The 2020 graduation rate for students who did not need DE at 2-year colleges was 1.5 times great than students who needed DE, whereas at 4-year colleges, two times as many students who did not need DE graduated than students who needed DE (THECB, n.d.-b, n.d.-c).

Although data reports provided through the annual Texas DE and persistence and graduation reports are quite informative, they include no indication of a unique number of students who enrolled in any DE course for any given year or time frame. Also, there is no gateway course success rate provided for students who enrolled in and passed DE. Finally, the reports lack any breakdown of student metrics by gender, race/ethnicity, age group, or other subpopulations (THECB, n.d.-a). However, ethnicity data on high school students who graduated and directly enrolled in Texas public colleges without meeting minimum college readiness scores on the TSIA is available in an annual report (THECB, 2021). For students who entered public IHE in Texas in fall 2019 directly after they graduated from a Texas high school. This population comprised approximately 77% of all first-time public college enrollments in fall 2019 (THECB, 2020a, 2020b, 2020c, 2021); therefore, a gross estimate of demographic composition is possible. Table 5 provides demographic information on the percentages of 2017 Texas seniors who did not meet college readiness levels on the TSIA prior to graduation. (TSIA scores may be substituted for high school end-of-course and high school exit assessments) [19 Tex. Admin. Code §101.4003], 2021).

Although I reported national data was for DE enrollment rates (see Table 3) and Texas data for below college-ready enrollment rates (see Table 5), a comparison of the differences in rates of students underprepared when entering college across race/ethnicity groups revealed that Texas has much greater gaps between groups than seen nationally. For example, there was almost a 20-point percentage difference between the academic readiness of White and Hispanic students, the two largest demographic groups, with White students generally entering college more academically prepared than Hispanic students. Notable in Texas DE participants was the very low percentage of Asian students who were not college-ready as opposed to the much higher percentage of Black or African American students; the gap between the college readiness of these two groups was over 40 percentage points.

A valuable insight into the composition of the students at the lowest levels of literacy who enrolled in DE in Texas comes from a joint research project of the RAND Corporation, the American Institutes for Research, and THECB. Daugherty et al. (2019a) designed the study to assess the efforts of Texas public colleges had made to "provide improved support to students who enroll in college and are assessed at the lowest levels of readiness" (p. iii). The authors compared demographic characteristics of students in Texas who placed into ABE levels 1 through 4 to students to students who placed as DE or college-ready in fall 2015. In Table 6, I replicated Daugherty et al.'s (2019b) descriptive statistics based on TSIA placement levels. There are striking differences between students placed in DE or college-level coursework, who tested as having

Table 5

Texas Fall 2019 College Readiness Levels by Student Characteristics for Students Enrolling in Public Postsecondary Directly After High School

	Below College Readiness Levels (%)		
Student Characteristic	Mathematics	Reading	Writing
All Students	34.8	20.5	10.8
Gender			
Female	36.1	20.5	9.5
Male	33.2	20.6	12.3
Ethnicity/Race			
Asian	8.6	8.0	3.4
Black or African American	50.7	32.3	17.5
Hispanic	41.7	26.1	13.1
American Indian or Native Alaskan	34.0	11.2	11.2
Hawaiian Native or Pacific Islander	35.2	30.1	14.5
More than one race	28.2	13.8	7.4
White	24.5	10.7	6.5
Economically disadvantaged	45.4	29.1	15.0

Note. Data sourced from the THECB *Annual TSI High School Summary Report: Student Performance on Texas Success Initiative (TSI) Readiness Measures 2018–2019 High School Graduates Enrolled in Texas Public Higher Education in Fall 2019* (THECB, 2021).

academic skills at ninth- to twelfth-grade levels, and students in ABE levels 1 through 4, whose academic skills were assessed below high school proficiency. Compared to students who placed into DE or college-ready levels, students placed into ABE levels were twice as likely to be Black or African American or older than 25, almost three times as likely to be of limited English proficiency, and over three times as likely to report disability.

Table 6

Student Demographic Characteristics by TSIA Placement Levels for First-Time Enrolled Students in Texas Public Colleges in Fall 2015

Student Characteristics	ABE Level 1–4	Developmental Education and College Ready
25 years or older	16.2	8.8
Black	22.4	12.1
White	16.6	30.3
Hispanic	52.1	48.7
Male	43.8	47.1
Limited English proficiency	9.2	3.7
Disability	6.6	2.0
Economically disadvantaged	65.4	50.9

Note. Statistics represent the percentage of first-time-in-college students with various characteristics by test score, comparing students with ABE level scores of 1–4 on the TSIA with students testing at higher levels. This table is from Daugherty et al.'s (2019b) "Supporting College Enrollees Who Test at the Lowest Levels of Readiness Lessons from Texas Community Colleges—Technical Appendix," p. 21.

Developmental Education: Research on Student Outcomes

Numerous researchers have reported on both short- and long-term academic outcomes for students who enrolled in college DE coursework. When comparing outcomes for students who participate in DE, it is important to recognize that regardless of participation in DE, first-generation college students and minority students are more likely to drop out of college than students who are white or who had one or more parents with some college education (Crisp & Delgado, 2014).

Traditionally, IHEs used a single assessment score to determine student placement into DE or college-level coursework. IHEs either required students or advised them to successfully complete one or more of several levels of DE reading, writing, and mathematics before they allowed them to enroll in entry college-level English or descriptive statistics based on TSIA placement levels. There are striking differences between students placed in DE or college-level coursework, with academic skills at mathematics. Over the past decade, many colleges have started to use measures in addition or instead of a single test score and to accelerate DE in order to reduce the number of levels students need to complete (Rutschow et al., 2019).

In order to compare outcomes for students with similar academic ability who enroll in DE versus those who enroll in college-level coursework, researchers may find postsecondary systems that allow a variety of cut scores to be used for placement purposes. In this way, researchers can compare the performance of students with the same placement test score who enroll in DE versus college-level gateway courses. Researchers have also used regression discontinuity design (RDD) to compare outcomes from students with similar characteristics and academic ability who enroll in DE as opposed to those who directly enroll in college-level coursework. In RDD, researchers compare outcomes for students at the margin of college readiness, both slightly above and slightly below, to determine if placement into DE had negative, null, or positive effects. Most researchers also use other control variables by including gender, ethnicity, race, and other student characteristic in addition to the student test scores into the prediction. The advantage of RDD is the ability to determine causal effects without using randomization, but the disadvantages include possible reduced statistical significance as opposed to randomized experimentation, possible untestable assumptions, and the limited interpretation of the results to only students who score close to the college readiness cutoff (Valentine, et al., 2017).

While several studies have shown positive outcomes for at least some DE subpopulations (Attewell et al., 2006; Bahr, 2010; Boatman & Long, 2018; Calcagno & Long, 2009; Lesik, 2007; Moss et al., 2014), other researchers have concluded that most students would have similar educational outcomes without the additional time and expense of DE classes (Bailey et al., 2010; Belfield & Crosta, 2012; Hodara et al., 2012; Jaggars & Stacey, 2014; Scott-Clayton et al., 2014). Several studies that compared enrollment patterns and success rates of students who tested slightly above cut-off scores for college readiness to those slightly below those cut-off scores (i.e., RDD) have findings of mostly null or adverse effects of DE placement on student grades, persistence, and college completion rates (Jaggars & Stacey, 2014; Martorell & McFarlin, 2011; Melguizo et al., 2016; Scott-Clayton & Rodriguez, 2015; Xu, 2016; Xu & Dadgar, 2018). These researchers have suggested that students who are close to being college-ready upon enrollment would have fared better if they had initially been placed into college-level classes.

Though findings from several studies included few positive outcomes for students who were initially assigned to, enrolled in, and participated in at least one DE course, the same is not true for research that examines outcomes for students who do complete their DE requirements. Researchers who focused on outcomes for students who completed required DE have concluded that students who completed their required DE coursework demonstrated increased retention (Fike & Fike, 2008), increased gateway math grades

(Moss et al., 2014), and similar or slightly higher rates of degree attainment (Noble & Sawyer, 2013; Schak, et al., 2017). For example, Fike and Fike (2008) investigated several possible predictors of in a community college system and discovered that the highest predictor for retention was passing a DE reading course, and passing a DE mathematics or writing class was also a significant predictor.

Noble and Sawyer (2013) studied differences between the degree attainment of students in 4-year IHEs who completed DE coursework and those who were college-ready at enrollment. Within the 6 years, fewer DE completers had earned their college degree than had college-ready students. However, when the time to completion was extended, the rate of degree completion between the two groups of students was similar. Moss et al. (2014) used an interesting research design that embedded a small, randomized experiment within a regressive discontinuity design to discover the effects of DE math participation on gateway course outcomes. Results from both the randomized experimental design and the regressive discontinuity analysis indicated that students who completed their required DE math sequence earned approximately one eighth to one fourth of a grade letter higher in their gateway math course than students with similar COMPASS test scores who directly enrolled into college-level gateway math.

Finally, Schak et al. (2017) recently asserted that national data indicate that the completion of developmental courses relates to increased persistence and degree completion in 2-year colleges. Nationwide data illustrates that for students who initially enrolled in 2-year IHEs in 2003, 26% who completed required DE coursework attained an associate degree within 6 years compared to 24% of students who were college-ready upon postsecondary entry (Schak, et al., 2017). Overall, research has suggested that students who completed their required or recommended DE courses have similar or slightly better academic achievements than students who did not participate in DE.

Postsecondary Assessment and Placement: National Practices

The definition of college readiness and subsequent referral to and placement into DE varies widely based on state or institution regulations. What constitutes college readiness typically involves the use of scores from a standardized academic skills test, such as high school exit exams, college admissions tests such as the SAT and ACT, or college placement tests such as the ACCUPLACER (Desai, 2012). Other factors that may be considered when determining academic college readiness include high school grade point average, the type of high school coursework completed, or exams designed by individual institutions (Whinnery & Pompelia, 2019).

A 2011 survey asked colleges about their current reforms in DE (Rutschow et al., 2011). Only four strategies out of a total of 40 involved assessment and placement. Rather than colleges making changes in assessment and placement, however, they were providing short-term instruction for improving placement test scores. Although effective for some students, this was not, in the pragmatic definition of the best solution to a problem, the most beneficial solution for all stakeholders. Almost a decade later, a second survey of the DE landscape included reforms in assessment and placement practices as well as "implementation of instructional and student support interventions designed to improve students' success" (Rutschow et al., 2019, p. E4).

Using Multiple Measures for Placement

A national trend over the past few years is the use of multiple measures for college placement. Rather than using a single college placement or college admissions test score, many states require some or all of their public higher education systems to factor in high school cumulative GPA or GPA for specific mathematics and English courses, and in a few cases, non-cognitive assessments (Hodges et al., 2020). The Texas Success Initiative Assessment (TSIA) explored in this investigation was itself a reform strategy for DE assessment and placement. Texas policy-makers sought to develop a placement test that was (a) aligned to the Texas College and Career Readiness Standards, (b) would provide diagnostic information for students whose scores were below the college readiness cut levels, and (c) would provide adult education levels that corresponding to those used in the federal adult education system. The diagnostics were to facilitate the development of DE interventions that offered instruction below ninth grade and the adult education levels were to indicate when students should be referred to adult education services rather than enrolled in DE interventions (THECB, 2014).

The definition of the term *multiple measures* differs across states, IHE systems, and researchers. Goudas (2019) contrasted the term multiple measures with what he defined as *single multiple measures*. Multiple measures occur when there is a simultaneous consideration of two or more measures for placement, whereas single multiple measures, which is more common, occurs when placement is based on one of several measures that include standardized test scores, high school overall or course-specific performance, non-cognitive assessments, or other factors. States and postsecondary systems may also employ tiered decision-making in which students must meet one of several multiple measures. For example, in North Carolina, students who do not meet the first combination of a minimum standardized test score and high school performance criterion may still qualify as college-ready if they meet criteria defined in a subsequent tier of combined test and high school performance (Hodges et al., 2020).

Prioritizing high school performance or using high school performance along with standardized test scores appears to be helpful in reducing the number of DE courses that students who are underprepared must take before they enter gateway college-level mathematics and English classes (Barnett et al., 2020; Hodara & Cox, 2016; Melguizo et al., 2016; Scott-Clayton et al., 2014). For example, results from two studies determined that assessment and placement using standardized test scores along with high school mathematics or overall GPA realized a significant increase for college-level mathematics placement without a significant decrease in college-level mathematics passing rates (Barnett et al., 2020; Ngo & Kwon, 2015).

However, reforms in math DE assessment and placement practices may be less effective for some DE students' sub-populations than for others. The Los Angeles Community College District noted that using an index of placement tests and HSGPA resulted in a lower percentage of African American and Hispanic students than White or Asian students into college-level coursework (Ngo & Kwon, 2015). In Florida, even after controlling for math skill levels using scores from Florida's placement test, researchers discovered that below-college ready students who were African American, female, or low-income chose more frequently than other student populations to skip DE math enrollment or to enroll in a prerequisite semester-long DE math course rather than to enroll in a corequisite model or to enroll directly into a gateway math course (Park et al., 2018). Barnett et al. (2020) reported that combining test scores with HSGPA increased college gateway course placement for all student subpopulations except for males in math, but also reported negative effects on gateway course grades for African Americans.

Using data analytics to determine customized assessment and placement policies at seven community colleges has shown promise. Researchers Bergman et al. (2021) worked with seven community college within a community college system that allowed colleges to create individualized placement procedures that led to an increase in placement into college English for all student demographic groups and into college math for all student demographic groups except for males. Furthermore, at most colleges, the gaps between student demographic groups in college-level placement decreased. Furthermore, students who were Pell recipients, African American, or females realized a significant increase in college credits earned. For multiple measures, all colleges used HSGPA and years since high school graduation and all but one college used at least one ACCUPLACER test score and GED status.

Using Optimal Cut Scores for Placement

Another emerging theme on improving the accuracy of college placement test accuracy that I identified was the need for institutions or institutional systems to set cut scores based on empirical data and to review the cut scores regularly. It is especially important to do so when students demographics or content standards upon which a test is based are updated (National Association for College Admission Counseling, 2016). Several researchers who analyzed the predictive validity of the COMPASS or ACCUPLACER scores on student course success determined that IHEs could improve placement accuracy rates while maintaining or gateway course grades by lowering cut scores (Barnett et al., 2020; Melguizo et al., 2013; Melguizo et al., 2016; Ngo & Melguizo, 2016).

Melguizo et al. (2013) outlined a process that IHEs could use to set optimal cut scores using empirical data. The authors also suggested that states with decentralized higher education government offer technical assistance to local IHEs to assist them in setting optimal cut scores. Camara et al. (2017) outlined methods to set college readiness cut scores using empirical data and IHE or IHE-system desired outcomes such as course grades or probability of passing a course. Belfield (2014) provided a practitioner toolkit with suggestions on using correlation analysis to check the strength of an IHE's current cut scores. THECB used the results of Cui and Bay's (2017) predictive validity study on the TSIA to revise the cut scores needed to demonstrate college readiness in writing (Morales-Vale & Montognese, 2017). Finally, the College Board offers a process for that assists IHEs in setting optimal cut scores for the ACCUPLACER (College Board, 2021).

After investigating the effects of policy changes in Florida on the accuracy of student placement, Leeds and Mokher (2020) determined by refining the cut scores, they were able to equal the placement accuracy of using multiple measures placement that considered both HSGPA and test scores from Florida's college placement test. The authors advised that states and large IHE systems use data analytics to determine optimal cutoff scores and to predict the accuracy of student placement using those cutoff scores prior to engaging in a large-scale effort to use HSGPA either in combination with college placement test scores or as part of complex decision tree scenarios for placement. In summary, to ensure that assessment and placement reforms benefit all students equally, educational systems must take care in their design, implementation, evaluation, and scale effective initiatives for assessment, placement, and instructional delivery content and format (Boylan & Bonham, 2011; Fong, et al., 2015).

Student Self-Placement

Researchers have largely focused on student self-placement into writing DE versus gateway English placement at 4-year institutions, and for the most part, have found models that follow Royer and Gilles (1998) directed self-placement process to be successful (Toth, 2019). However, the same has not been found at community colleges. Researchers examining the effects of Florida's self-placement policy reported an increase in the percentage of students who enrolled in college-level coursework and a decrease in the percentage of students who enrolled in DE. Although more students passed college gateway courses because more students enrolled in them, the failure rate in college gateway courses has increased (Hu et al., 2016). Researchers examining self-placements for students who would have been required to enroll in DE math prior to Florida's legislative change discovered that about 35% enrolled in DE math, about 35% enrolled in no math, almost 30% enrolled in gateway math, and less than 5% enrolled in a corequisite DE math model. After controlling for academic ability based on standardized test scores, students who were female or African American were the most likely to enroll in DE rather than in gateway or corequisite DE and also the groups most likely to skip enrollment in DE or college-level math (Park et al., 2018).

A recent study examining the effects of self-placement into math at one of nine colleges in the Los Angeles Unified Community College District that used self-placement for DE math during the 2008–2009 academic year (Kosiewicz & Ngo, 2020). Students were required to meet with advisors, but had no COMPASS scores available as part of the advising process. Compared to a control group placed using COMPASS scores, female, Black, White, and Asian students were less likely to withdraw from their DE math course. However, only male, White, and Asian students were more likely to pass a college level math class within 4 years. A second study by Coleman and Smith (2021) in a community college setting found that some were pressured by faculty advisors to enroll in DE, and in general, faculty underestimated students' likelihood of passing their courses; the authors reported that "47% of the English students and 29% of the math students who faculty felt were placed too high earned a C or better, and in English most of those students actually earned an A or B" (p. 417).

Park et al. (2018), Kosiewicz and Ngo (2020), and Coleman and Smith (2021) all suggested that there may have been implicit bias on the part of advisors regarding student math ability based on gender, race, or ethnicity that affected advisors' recommendations for placement into DE despite students' self-directed placement results. However, it may also have been possible that students who were female or African American were more likely to have underestimated their math abilities and chosen to enroll in lower-level math courses despite self-placement results that indicated they were likely to pass gateway courses. Regardless of the reason for heterogeneous effects, these researchers concluded that more research is needed on self-directed placement in order to ensure that it does not advantage some student populations more than others.

Postsecondary Assessment and Placement in Texas

Texas created its first statewide college placement initiative in 1989 with the development of the Texas Academic Skills Program (TASP). Per McAdams (2017), "as an assessment and a policy, the TASP was the State's attempt to unify testing practices at public institutions, thus ensuring students have the basic academic skills to complete successfully college level courses" (p. 2). TASP administration was required for students who could not show college academic readiness through other types of approved standardized tests. Students who did not meet the college readiness cut scores on the TASP were required to take DE classes and then to retake the TASP before moving into college-level coursework.

Most likely in an effort to differentiate the TASP initiative from the TASP assessment, the TASP assessment was renamed to the Texas Higher Education Assessment (THEA) soon after its implementation. In 1993, test developers added diagnostic information to the THEA's student score report to assist with placement decisions (Pearson Education, 2013). Although not used currently in Texas public higher education, the THEA is still in use by some private postsecondary institutions (McAdams, 2017). In 2003, the 78th Texas Legislature replaced the Texas Academic Skills Program with the Texas Success Initiative (TSI). The 2003 TSI allowed a wide array of approved placement tests (e.g., the THEA, ASSET, COMPASS, or ACCUPLACER), and institutions could increase the minimum cut-off scores set by the state (McAdams, 2017).

With little evidence of increasing college success rates for all Texans, THECB recommended to the 82nd Texas Legislature the development of a single college placement instrument, the TSIA, that would be central to a comprehensive assessment and placement policy. Texas approached the development of the TSIA with the intent of providing a process for assessment that would circumvent some of the typical problems that affect issues of predictive validity and placement (THECB, 2014). Before attempting the TSIA assessment, a student had to participate in a pre-assessment activity that addresses the importance of the assessment to the student's college experience, an explanation of the testing process and practice test questions, development education options, and student resources. Finally, Texas institutions must use a holistic placement for students who are not exempt from the TSIA and whose scores are below the cut-off for college readiness. Multiple measures that colleges must consider for placement include the student's HSGPA and rank, former school and work experiences, personality traits such as motivation, and personal life issues that may either support or hinder the student’s progress (19 Tex. Admin. Code §4.55).

As of fall 2013, the TSIA became the sole college placement test for use in IHEs in Texas (19 Tex. Admin. Code §4.56). A version of the ACCUPLACER developed specifically for Texas, the three TSIA tests–reading comprehension, mathematics, and writing–classify students as either having college level, high school level, or below high school level academic skills (THECB, 2015). The state legislature sets the cut scores for college readiness upon recommendation from THECB, and IHEs may not alter them (19 Tex. Admin. Code §4.57). Several exemptions exist to the TSIA assessment, including satisfactory scores on college admissions tests (ACT and SAT), on Texas high school exit exams, prior college credits earned, military service, waivers for non-native English speakers, enrollment in certificate programs requiring less than one year and 42 hours for completion, and participation in other institution-based interventions (19 Tex. Admin. Code §4.54). On January 11, 2021, the THECB replaced the TSIA with the TSIA 2.0; however, TSIA scores and TSIA ABE Diagnostic placements are valid for up to 5 years after the student’s test administration date (19 Tex. Admin. Code §4.56).

To better place and serve academically students who are underprepared, Texas policymakers designed the TSIA to function as both a placement test and, for students who score below the college readiness cut-off, a diagnostics assessment. After the initial college readiness placement testing is complete, students who do not score at college-ready levels immediately receive either a DE diagnostics assessment or an ABE Diagnostics and placement assessment. Performance on the ABE Diagnostic assessment determines the student's placement into DE at the college or, for students with very low skills, referral to FF-ABE services (THECB, 2015). Diagnostic information on the ABE Diagnostic score report includes proficiency ratings for four content areas in each subject area (College Board, 2018b). Students who take the ABE Diagnostic test also receive a numeric placement score between 1 to 6. These scores indicate the ABE educational functioning level of the student as described by the federal academic measurement scale developed by the NRS. The NRS (2021) uses an educational framework that divides adult basic skills into six educational functioning levels that correspond to grade-level equivalencies as follows:

- NRS level 1 corresponds to grades K–1 equivalency
- NRS level 2 corresponds to grades 2–3 equivalency
- NRS level 3 corresponds to grades 4–5 equivalency
- NRS level 4 corresponds to grades 6–8 equivalency
- NRS level 5 corresponds to grades 9–10 equivalency
- NRS level 6 corresponds to grades 11–12

The THECB requires students who test below college readiness cut-off scores on the TSIA to be placed into a variety of DE options or basic adult skills education non-course competency-based developmental education options (NCBOs), which are non-traditional interventions designed to accelerate the remediation needed to reach desired proficiencies (Morales-Vale & Daniels, 2015). Beginning in the 2018–2019 academic year, Texas IHEs were required to enroll at least 25% of students who scored into ABE levels 5 or 6 into a DE corequisite model defined as follows: “An instructional strategy whereby undergraduate students…are co-enrolled or concurrently enrolled in a developmental education course or NCBO . . . and the entry-level freshman course of the same subject matter within the same semester” (19 Tex. Admin. Code §4.53[7]).

For the 2021–2022 academic year, 100% all students who place into DE levels must enroll in corequisite models (19 Tex. Admin. Code §4.62[8]). However, college advisors will continue to have discretion for the placement or referral of students who score an ABE level of a 1 through a 4 on the TSIA diagnostic. The THECB recommended that students with ABE levels of 3 or 4 be enrolled in a basic adult skills education (BASE) NCBO. THECB also recommended that students with TSIA diagnostic levels of 1 or 2 in all three subject areas be referred to FF-ABE services (Morales-Vale & Daniels, 2015). See Figure 2 for an illustration of the path through the mathematics TSIA along with required and allowable placements based on the student’s college ready, DE, or ABE levels.

Fairly unique to Texas is a legislative mandate for holistic advising for students who place under college readiness levels on the TSIA. Daugherty et al. (2021) used student data from Texas public colleges between fall 2013 and spring 2015 to investigate “student characteristics that are related to academic preparation or might signal a student’s likelihood of success or need for additional support and might therefore be

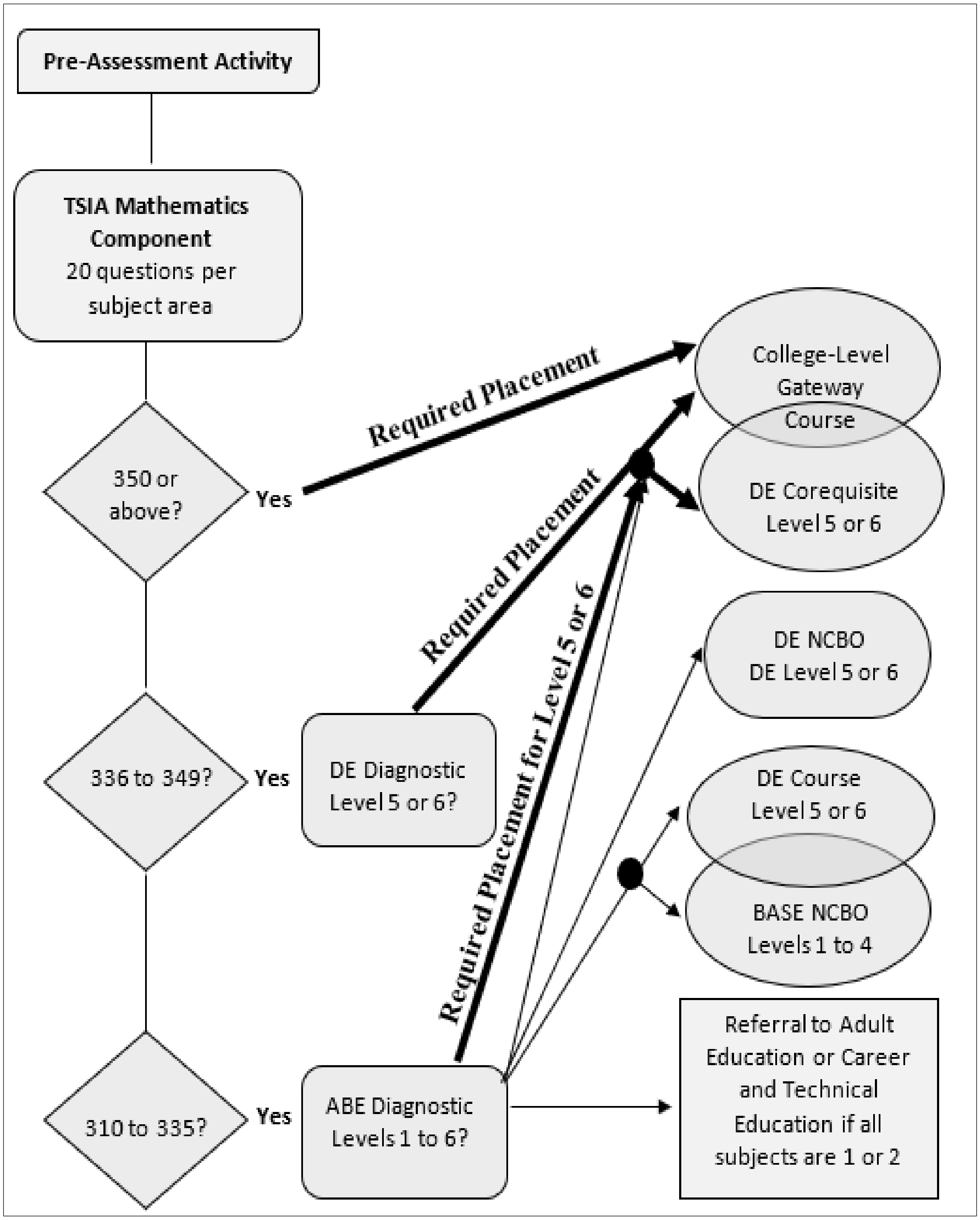

Figure 2. Path Through the TSIA and Possible Placements or Referral as of Fall 2021**.** This Illustration was adapted from "College Readiness and TSI Updates" by S. Morales-Vale and K. Montognese (Morales-Vale & Montognese, 2017, p. 21).

factors considered for placement into college-level courses under ‘holistic advising’ or ‘multiple measures’ initiatives” (p. 1086). The researchers used a regression discontinuity design to study a narrow band of students near the cut scores for college readiness. The only student characteristics found to have a negative impact on the probability of gateway course success were part-time enrollment and being over the age of 21. Student characteristics found to help predict gateway course success, some of which may seem counterintuitive, were the following: (a) scoring just under the minimum cut scores for college readiness, (b) having limited English language proficiency, (c) being low-income, (d) declaring a major (e) pursuing a bachelor's degree, and (f) attending full-time. Results from a replication of this study with more recent student data from Texas community colleges might assist advisors with the placement of students who score just below DE readiness level into corequisite DE, stand-alone DE coursework, or DE or ABE non-course-based options.

The purpose of the TSIA assessment system was placement. Scores from the TSIA, placements from the TSIA ABE Diagnostics, and subject specific proficiency ratings served the purpose of measuring student proficiency as well as providing additional diagnostic information for a more informed and accurate placement system (THECB, 2014). To document evidence of positive change in placement practices for students with TSIA NRS levels between 1 and 4, the THECB annually surveyed public IHEs about their placement practices based on student’s TSIA levels. In 2015, when asked if their advising practices, rubrics, or guidelines were differentiated for students who tested at NRS Levels 1-4, 72% of Texas public institutions responded affirmatively (THECB, 2016).

The 2015, 2016, and 2017 DE program surveys included questions asking about college's most common form of placement of students who scored at NRS Levels 1-4 (Daugherty et al., 2019b; THECB, 2016, 2018a). As illustrated in Table 7, the most common methods for placement or referral of students with TSIA ABE Diagnostic placements levels of 1 to 4 shifted. The largest increase over the 3-year period was in student referral to FF-AEL services. Although placement into DE courses with concurrent adult basic skills education (BASE) support increased, students were still referred to DE courses without BASE support more often than DE with support. One of the reasons for this was due to the small number of students at many colleges who tested at TSIA ABE levels between 1 and 4, making it difficult for classes to have a minimum number of students necessary. Another reason was because there was no mandate from the state for students to receive additional BASE support in their DE classes, many colleges did not require students to opt for the additional support. However, over the 3-year period, the shift to referral to FF-AEL services for these lower-skilled students and placement into DE courses with additional BASE support is evidence of an increased awareness by college leadership and advising staff of differentiated placement strategies that could better meet the individual needs of these students (Daugherty et al., 2019a).

Adult Basic Education

I now transition from my review of DE to present literature related to federally-funded adult basic education (FF-ABE) services. Regulations of the National Reporting System for Adult Education (NRS) dictate most aspects of FF-ABE programming and virtually all assessment and placement requirements for FF-ABE students. Historically, ABE has been viewed as a program to ensure that the US economy remains healthy

Table 7

Most Commonly Used Placements or Referrals for Students with NRS Levels 1 through 4 in Texas Public Postsecondary Education

Most Common Placements or Referrals	Year		
	2015 [a]	2016 [b]	2018 [c]
DE courses with BASE NCBO	25	35	37
DE courses without BASE NCBO	35	28	42
TSIA test and re-test preparation programs	17	20	11
Referral to Adult Education and Literacy (AEL) programs	6	4	25
Career and technical education (CTE) courses with BASE NCBO	1	2	0
Career and technical education (CTE) courses without BASE NCBO	3	5	8
Not applicable or no response	13	6	N/A[c]

Note. [a] Data from *Developmental education: Updates and progress for underprepared students, 84th Texas Legislature* (THECB, 2016, p. 24). [b] Data from *Supporting college enrollees who test at the lowest levels of readiness: Lessons from Texas community colleges* (Daugherty et al., 2019a, p. 14). [c] Data from *Developmental education: Updates and progress for underprepared students, 85th Texas Legislature.* The questionnaire allowed multiple responses, so the total is over 100% (THECB, 2018c, p. 31).

healthy and that adults have the necessary training to fill jobs within society. The educational focus is workforce training rather than higher education, so ABE program outcomes focus on entry into the job force rather than long-term educational attainment (Jacobs & Tolbert-Bynum, 2008).

The demographics of the students served vary by state, as do the amount and type of required state match and the state-level fiscal agent. Approximately 50% of states run ABE through the community college system, whereas others manage ABE through the state's workforce or K–12 systems (Jacobs & Tolbert-Bynum, 2008). Following is a national and Texas-specific synopsis of the history and current landscape of FF-ABE services, participants, assessment and placement practices, and student transitions to postsecondary education.

Adult Basic Education Facts and Figures

Adult Basic Education (ABE), as we know it today, began in 1964 when the federal Economic Opportunity Act established the ABE Program and encouraged states to work in partnership with the federal government to expand educational services for adults with literacy needs. In 1991, the National Literacy Act provided a much more comprehensive ABE system, including the establishment of the NRS for documentation of accountability. The 1998 Workforce Investment Act encouraged partnerships between workforce, education, and community resources. Title II of this act, the Adult Education and Family Literacy Act, addressed services for adult basic skills education (U.S. Department of Education, 2013). Recently, the national emphasis on the need for skilled labor that meets the demand of the twenty-first century spurred the development of the newly enacted Workforce Investment and Opportunity Act (WIOA). WIOA "emphasizes workforce preparation and postsecondary education as the 'core purpose' of federally-funded ABE/ASE programs" (Pickard, 2016, p. 50).

Participation in federally funded ABE (FF-ABE) services increased from an initial enrollment of 37,991 students in 1965 to a high of over 4 million students in 1995

(Office of Vocational and Adult Education, 2013). Since 1995, the student population served has before gradually decreased. For the 2019–2020 program year, the federal government awarded approximately 6 million dollars (Office of Career and Technical Education, 2021) to help states serve the close to 1.1 million adults who participated in FF-ABE services (NRS, 2020g).

During the 2019–2020 FF-ABE program year, 51% of adults participated in basic skills education and 49% in English language instruction. Gender was distributed almost equally, with 49% female and 51% male. The most populous ethnic or racial group was Hispanic at 46%, and half of all students fell in the 25 to 44 years age range (NRS, 2020c). Over one fourth of all adult basic education students entered FF-ABE programming at NRS levels 2 and 3, which corresponds to basic academic skills at second- through fifth-grade levels (NRS, 2020a). See Tables 8 and 9 for additional demographic statistics.

Compared to the national FF-ABE system, for the 2019–2020 program year, Texas demographic characteristics were somewhat similar to national characteristics for age and the distribution of students in most NRS level placements. However, program participation by gender differed strikingly; nationally, participation was almost equal for males and females, whereas in Texas, participation was 71% female and 29% male. At 73.3%, Hispanic enrollment in Texas adult education is also much higher than the national Hispanic enrollment of 46.6% (NRS, 2020c, 2020d). In ESL programming, Texas had a much higher percentage of students who entered ESL at the lower two levels, approximately 32% as compared to the 14% at the national level. Participation in Texas

Table 8

2018–2019 Nationwide and Texas ABE Participant Demographic Characteristics

Participant Demographic Characteristics	Nationwide [a] n = 1,280,540 %	Texas [b] n = 78,047 %
Sex		
Male	41.1	29.0
Female	48.9	71.0
Age		
16–18	8.4	6.0
19–24	17.6	15.3
25–44	49.0	51.7
45–54	14.7	17.1
55–59	4.6	4.8
60+	5.9	5.1
Race or Ethnicity		
American Indian or Alaskan Native	1.0	0.3
Asian	10.0	5.7
Black or African American	17.6	9.8
Hispanic or Latino	46.6	73.4
Native Hawaiian or Other Pacific Islander	0.5	0.1
White	22.3	10.1
More than One Race	2.1	0.6

Note. [a] Percentages calculated from online data (NRS, 2020c). [b] Percentages calculated from online data (NRS, 2020d).

of advanced ESL students was only 1.6% compared to 6.9% nationwide, with only 3% of academic and English language skills than does the nation overall (NRS, 2020a, 2020b). See Tables 8 and 9 for a more detailed breakdown of national and Texas FF-ABE student demographics and literacy levels.

Table 9

2019–2020 Nationwide and Texas ABE Participant NRS Levels at Enrollment

Participant Enrollment NRS Level	Percent [a] Nationwide n = 1,280,540	Percent [b] Texas n= 78,047
Program Enrollment		
Adult Basic Education	41.0	38.4
Adult Secondary Education	7.3	4.3
English as a Second Language	51.7	57.4
Educational Functioning Level		
Adult Basic Literacy	3.2	2.3
Beginning Adult Basic Education	13.3	13.2
Low Intermediate Adult Basic Education	14.3	13.6
High Intermediate Adult Basic Education	10.2	9.3
Low Adult Secondary Education	4.4	3.2
High Adult Secondary Education	3.0	1.1
Beginning ESL Literacy	6.6	15.3
Low Beginning ESL	7.1	16.6
High Beginning ESL	10.1	10.4
Low Intermediate ESL	11.4	6.1
High Intermediate ESL	9.6	7.4
Advanced ESL	6.9	1.6

Note. [a] Percentages were calculated from online data from the NRS for Adult Education (NRS, 2020a). [b] Percentages were calculated from online data from the NRS for Adult Education (NRS, 2020b).

Up until 2012, the Texas Education Agency received the federal block grant from the U.S. Department of Education. At that time, Texas policy makers expressed their desire to increase the alignment of FF-ABE services with further education and job training and passed Texas Senate Bill 307, which transferred ABE to the Texas Workforce Commission (TWC) on January 13, 2013 (Sunset Advisory Commission, 2013, 2014). Texas experienced an initial drop in participants served, going from a 2012 total of 91,906 (NRS, 2012b) to 71,676 in 2013 (NRS, 2013). Enrollment peaked in the 2015 program year with 102,083 participants (NRS, 2015a) but dropped to 70,018 by the 2019–2020 program year, with 57.4% enrolled in ESL and 42.6% enrolled in ABE (NRS, 2020c).

Adult Basic Education Assessment and Placement

Federal NRS policies mandate baseline and progress assessment of adult learners enrolled in FF-ABE programs through federally approved assessments. These assessments must align with the NRS level descriptors. These descriptors, which align with the U.S. Department of Education's *College Readiness Standards for Adult Education* (CCR-AE), describe the expected student outcomes upon completion of each NRS level in three competency areas: basic reading and writing, numeracy skills, & functional and workplace skills (OCTAE, 2016).

In FF-ABE programs, all assessment must occur before the student begins instruction and must re-occur after a minimum number of instructional hours, which varies based on the assessment and the student's baseline test scores. Currently, there are two approved national assessments for students who enter the FF-ABE system with the desire to improve basic skills in reading, writing, and mathematics. These include the TABE 11/12 and the Comprehensive Adult Student Assessment System CASAS. (NRS, 2021). Assessment practices for Texas FF-ABE services must abide by federal requirements. States do, however, have the option of choosing which of the approved tests they will use for assessment. Texas currently allows either the TABE 11/12 or the CASAS for adult basic and secondary education students. Students must baseline and progress using the same assessment, and students who test before the test publisher's recommended minimum hours of instruction must have documentation on file to demonstrate their academic readiness to progress test (Texas Workforce Commission, 2020).

Adult Basic Education Postsecondary Transition

U.S. Department of Education defines FF-ABE participant outcomes through its National Reporting System for Adult Education (NRS). Beginning in 2000, the NRS collected data on postsecondary outcomes, but the methodology changed drastically over time. Tracking began based on adult students' stated goals but then changed to a pre-defined cohort of students that was based on the student's NRS level at the beginning of the program year (NRS, 2012a). Between the 2000–2001 and 2011–2012 FF-ABE program years, the NRS required states to report on skills gains based on pre-and post-testing with an approved FF-ABE assessment (such as the TABE 11/12). The NRS also required states to report the number of HSE certificates attained by participants during their FF-ABE programming and if participants who exited from FF-ABE services obtained employment. The final outcome measure was postsecondary enrollment. Unfortunately, an examination of the methodology used to calculate the postsecondary enrollment measure revealed that this outcome only pertained to participants who had a documented goal of college transition by the end of the FF-ABE program year. Therefore, the outcome measured the attainment of a participant's personal goal to enter postsecondary education rather than the actual postsecondary enrollment of FF-ABE students following their exit from FF-ABE services (NRS, 2012a)

Beginning in the 2012–2013 program year and continuing for 4 years, the NRS changed its methodology for collecting postsecondary enrollment data by tracking postsecondary transition for FF-ABE participants who exited during both the current and prior program years. The NRS identified the postsecondary transition cohort consisting of all participants who had entered the program already holding a secondary credential, who attained a secondary credential during the program year, or who had enrolled in a class designated as a college transition class (NRS, 2012a). Nationally, 29% of the 2015–2016 postsecondary transition cohort and 30% of the 2014–2015 postsecondary transition cohort entered postsecondary education during the 2015–2016 program year (NRS, 2016a), making a total transition rate of 59% for the 2014 and 2015 cohorts. In Texas, postsecondary transition rates for the same time period were slightly lower; almost 20% of the 2015–2016 postsecondary transition cohort and close to 25% of the 2014–2015 postsecondary transition cohort entered postsecondary education during the 2015–2016 program year (NRS, 2016b).

Postsecondary transition rates for students who exit FF-ABE programs with an HSE certificate may be higher than the general population of HSE certificate holders. Jacobs and Tolbert-Bynum (2008) estimated that approximately 30% of GED graduates are current or past FF-ABE participants. The GED Testing Service tracked its GED 2014 graduates over the same 2-year time frame used by the NRS for FF-ABE participants who exited in 2014 with HSEs. By the end of the 2015–2016 academic year, 41% of GED 2014 graduates had enrolled in postsecondary (GED Testing Service, 2017), which is a lower transition rate than the 58% reported by the NRS for FF-ABE participants over the same time period (NRS, 2015b).

With the passage of WIOA, the NRS required states to report several types of data concerning educational progress. One type of educational gain, considered an educational functioning level gain, occurs when in the same program year, a student enrolled in basic FF-ABE services enrolls in postsecondary education or training after exiting FF-ABE services (NRS, 2021). The student's postsecondary enrollment must be into at least one course taught at postsecondary skill levels; therefore, a student's enrollment into only developmental education or ESL coursework is not counted toward postsecondary transition (Green, 2019). Nationally, approximately 10% of all adult basic education students who exited adult education services during the 2020 program year enrolled in postsecondary education during the same year (NRS, 2020e). However, when looking only at transition rates for students who started the year with skills levels of at least 6th grade (NRS level 4 or above), the postsecondary transition rate was almost 60%. In Texas, the transition rate for all adult basic education students was only 3.6%, and for students who began the year with at least 6th-grade skill levels, the rate was approximately 30% (NRS, 2020f).

WIOA also requires states to report on the attainment of postsecondary education and training credentials when they occur while the student is still enrolled in FF-ABE services and up to the end of the program year that follows the student's exit from services (NRS, 2021). In contrast to the nation's higher percentage of FF-ABE students who enroll in postsecondary education, Texas realized higher attainment of postsecondary credentials for students who are still enrolled in FF-ABE or within a year of their exit. For this outcome measure, Texas realized a rate of close to 38% (NRS, 2020f), whereas nationally, the rate was approximately 20% (NRS, 2020g). These higher numbers for credential attainment numbers might represent Texas Workforce Commission (2020) policies that have required an increasing number of students to enroll in a WIOA program model called Integrated Education and Training. This model co-enrolls students in adult education basic skills classes along with one or more postsecondary courses. Although many of these postsecondary courses are non-credit, they must lead to the attainment of any higher education certificate or degree.

However, the higher Texas outcome might be attributed to the variety of processes that states use to collect students' post-exit enrollment into postsecondary education or training. For example, when the NRS began tracking postsecondary enrollment in Texas for the 2014–2015 program year, the Texas Workforce Commission (TWC) used data available from the Texas Higher Education Coordinating Board to report a 2014 transition rate of 10%. According to the TWC ABE Director at that time, these low transition rates were mostly attributable to this data matching process because the THECB did not require IHEs to report enrollment in or credential attainment from non-credit CTE classes and programs. Because many ABE students often choose non-credit certificate pathways, it was highly probable that there were many unreported student transitions. For this reason, data matching was discontinued for credentials awarded as programs were directed to document postsecondary credential awards for currently enrolled students and to follow-up with students after their exit during the subsequent program year to document any credentials they earn (A. Green, personal communication, August 1, 2018). Using TWC-monitored tracking and documentation at the local program level, postsecondary transition rates doubled from 10% in 2014 to 20% in 2015 (A. Green, personal communication, August 1, 2018). Inconsistencies between information collected THECB and TWC may be alleviated in the future by the recent passage of the Texas Education and Workforce Alignment Act that created a Tri-Agency Workforce Initiative between the Texas Workforce Commission, The Texas Higher Education Coordinating Board, and the Texas Education Agency for “sharing and cooperatively managing education and workforce information collected” (Texas Education and Workforce Alignment Act, §2308A.003(1), 2021).

Chapter Summary

In this chapter, I presented a review of the literature related to my research process of inquiry. According to Dewey's (1938) theory of inquiry, my time spent in the literature review was part of the problematic situation in which evidence is gathered in efforts to test the identified hypothesis or resolve an identified problem. I also approached the review through the lens of the *Standards* test validation framework that I adapted from AERA et al. (2014) *Standards for Educational and Psychological Tests* (see Chapter 1).

I began the reviewing with a review and synthesis of research on the use and validity of standardized tests to determine placement into college coursework or, in the case of Texas, referral to adult education services. I located much more research and evidence of validity for the SAT and ACT college admissions tests than for college placement tests (Bettinger et al., 2011; Sackett et al., 2009; Westrick et al., 2019). Test publishers ACT and College Board both recommended using college admissions and college placement test scores along with HSGPA to correct variance between demographic groups and to improve the predictive ability for student college success

(Bettinger et al., 2011; Mattern & Packman, 2009; Sackett et al., 2009; Westrick et al., 2014, 2019).

Next, I located research on the predictive ability for three commonly used college placement tests: the ACCUPLACER, the COMPASS, and the TSIA. Publisher-sponsored researchers for these three placement tests reported that overall placement accuracy for the math sections ranged from 63% to 67% (Cui & Bay, 2017; Mattern & Packman, 2009; Westrick & Allen, 2014). Independent researchers generally calculated similar placement rates for the ACCUPLACER and COMPASS (Belfield & Crosta, 2012; Scott-Clayton, 2012; Scott-Clayton et al., 2014). Many researchers reported that the reading and writing placement tests typically underplaced a larger percentage of students, ranging from 25% to 44%, than did the math tests (Barnett et al., 2020; Belfield & Crosta, 2012; Scott-Clayton, 2012; Scott-Clayton et al., 2014). Several researchers also reported that HSGPA tended to better predict grades in DE and in college-level gateway courses better than did ACCUPLACER or COMPASS test scores but that combining test scores with HSGPA proved to be the most accurate predictor for almost all tests and student demographic groups (Barnett et al., 2020; Belfield & Crosta, 2012; Ngo & Kwon, 2015; Scott-Clayton et al., 2014; Westrick & Allen, 2014).

Background knowledge and viewing every situation within its context are both very important tenants in pragmatic thought. In line with this pragmatic worldview, I next provided the reader with a broad overview of developmental education participation both nationally and in Texas. Because my study examined the relationships of math scores, I estimated demographic information for DE math participation in Texas to compare Texas and national student characteristics. Understanding group differences in assessment performance is a necessary component of my *Standards* test validation framework, as it falls into the areas of test fairness and consequences of test use.

Following my descriptive statistics on the prevalence of DE and the characteristics of students who enrolled in DE math, I presented recent research on DE student outcomes in which many researchers concluded that students who began in DE achieved reduced academic success outcomes compared to students who began in college-level coursework in large part due to failure to enroll in or to complete their DE courses. Students who completed required DE, however, appeared to perform as well as or slightly better than those who never participated in DE. The issue of student attrition from DE spearheaded many reforms over the past 10–15 years. Therefore, I researched current DE reforms, which I broadly categorized as reforms with the intent to accelerate DE, improve DE student support, and provide more accurate DE assessment and placement.

Because my investigation focused on assessment used for placement of students with the lowest academic proficiencies, I described the assessment and placement process generally used nationwide and specifically used in Texas public colleges. I also provided an overview of current reforms for DE assessment and placement, which included review of literature on the using multiple measures for college placement (Barnett et al., 2020; Belfield & Crosta, 2012; Ngo & Kwon, 2015; Scott-Clayton et al., 2014; Westrick & Allen, 2014) and research on the calculation and adjustment of cut scores used for placement (Morgan & Michaelides, 2005; Scott-Clayton et al., 2014). Finally, I included information on research that examined the effectiveness of alternative college placement systems that were found to be positive for most students but had unintended negative consequences for one or more subpopulations based on demographic characteristics (Barnett et al., 2020; Park et al., 2018), including self-placement (Coleman & Smith, 2021; Kosiewicz & Ngo, 2020; Park et al., 2018). Results from these studies on alternative placement models reinforce the fact that it is not the assessment but the way it is used for placement that must be validated to ensure that all students realize similar benefits.

III: METHODOLOGY

The purpose of this study was to compare student performance on a college placement test, the Texas Success Initiative assessment in mathematics (TSIA-M) and its adult basic education (ABE) diagnostic component, and an adult education test, the Test of Adult Basic Education in mathematics (TABE-M). I conducted the study using a conceptual framework of test validation that I adapted from the AERA et al. (2014) *Standards for Educational and Psychological Testing (Standards*). The study arose from a real-world context in Texas in which students' ABE placement levels from the TSIA-M and the TABE-M differed. Because both tests aligned their scores to the same measure—the National Reporting System's adult education functioning levels, I expected the placement levels from the two tests to be somewhat similar. College advisors use TSIA-M ABE Diagnostic levels to determine student placement into college-level or developmental education (DE) and may optionally refer students with ABE levels between 1 and 4 to adult basic education services. The results of this investigation will provide increased knowledge about the relationships between scores and placement levels from the two assessments so that students, advisors, and educators can make informed decisions that best support the student's journey through postsecondary education. In this chapter, I include details and discussion on the research design, IRB process, study setting, participant data set, instrumentation, variables, and the data analysis procedures I had selected to answer my research questions.

Research Design

I employed a non-experimental correlational research design for this study, for which I used correlation and a consistency of classification analysis to explore the relationships between TSIA-M scores, TSIA-M ABE Diagnostic placements, and TABE-M scores and placements. The research design I have chosen is typical for studies that explore evidence for test validation using relationships to other criteria. Although most research on the TSIA has focused on predictive validity, I performed a correlational analysis of the relationships between student performance on the TSIA and the TSIA ABE Diagnostic and the TABE 11/12 to search for evidence of convergent validity. Per Carlson and Herdman (2010): "Convergent validity reflects the extent to which two measures capture a common construct . . . [and] a form of evidence used to judge the construct validity of a measure" (p. 18).

I also analyzed relationships between performance on the two tests by age, gender, and race/ethnicity. By focusing on subgroups within a population, I searched for the presence of any unintended consequences of the test use based on student demographic characteristics. The *Standards* test validation framework I adapted from the AERA et al.'s (2014) *Standards in Educational and Psychological Testing* (see Chapter 1) guided me in the methods I chose to answer my research questions so that my research could both provide a practical solution to my inquiry as well as provide possible sources of evidence for validation of the TSIA-M and TSIA-M ABE Diagnostic assessments.

Research Questions and Null Hypotheses

The purpose of this investigation is to compare student performance on closely administered TSIA-M, TSIA-ABE Diagnostic, and TABE-M assessments by analyzing the relationships between student scores and placement levels on these instruments. Below are the research questions I formulated:

1. What is the relationship between student scores on closely administered TSIA-M

and TABE-M assessments?

2. What is the relationship between student NRS placements on closely administered mathematics TSIA-M ABE Diagnostic and TABE-M assessments?
3. Does controlling for age, gender, race, or ethnicity change the relationships between student scores on closely administered mathematics TSIA-M and TABE-M assessments?
4. Does controlling for age, gender, race, or ethnicity change the relationships between student NRS placements on closely administered mathematics TSIA-M ABE Diagnostic and TABE-M assessments?

To answer the above research questions, I tested the following null hypotheses:

1. There is no statistically significant relationship between student scores on closely administered mathematics TSIA-M and TABE-M assessments.
2. There is no statistically significant relationship between student NRS placements from closely administered mathematics TSIA-M ABE Diagnostic and TABE-M assessments.
3. There are no statistically significant differences in the relationships between student scores on closely administered mathematics TSIA-M and TABE-M assessments after controlling for age, gender, race, or ethnicity.
4. There are no statistically significant differences in the relationships between student NRS placements on closely administered mathematics TSIA-M ABE Diagnostic and TABE 11/12 assessments after controlling for age, gender, race, or ethnicity.

IRB Process

At the community college from which I requested data, external researchers must seek approval from their sponsoring institution prior to requesting data from the college. Therefore, I submitted my IRB application first to Texas State University, and in December 2019, I received approval and a human-subject exemption letter from Texas State University. The response was delayed due to the COVID-related college closure of facilities and turnover in leadership at the institutional research center. In November 2020, I was asked to provide the research site with an updated human-exemption letter from Texas State University. I received the updated human-exemption letter in December 2020 and sent it to the research site. The research institution approved my request in January 2021 and worked with me to provide the data I requested, which I received in April 2021.

Local Setting

The study's data comes from a medium community college (MCC) located in a suburban area in southeastern Texas. Due to confidentiality, the school requested not to have its name or any of its faculty or staff's names disclosed. The data collected for this study spans the period in which the TSIA-M and TABE-M form 11 or 12 were both approved for administration, which was from July 1, 2019, through January 11, 2021. In fall 2019, MCC's total enrollment (not including high school dual-credit students) was approximately 3,000 students, composed of approximately 60% females and 40% males; 20% Black, 35% Hispanic, 40% White, and 5% other; 60% ages 8 to 24; 30% ages 25 to 39, and 10% ages 40 and up. Also, in fall 2019, approximately 650 first-year students enrolled, of which about half scored below college readiness in mathematics on the TSIA (sources not revealed for confidentiality purposes).

Study participants were administered the TSIA-M, the TSIA-M ABE Diagnostic, and the TABE-M as part of the college enrollment and placement process at the study site. According to the adult education director at the community college, the college's philosophy is to use intrusive advising to direct students who place into ABE levels 1 through 4 on the TSIA ABE Diagnostic to enroll in an 8-week college success academy (CSA). With intrusive advising, "also known as proactive or engaged advising . . . staff take the initiative in reaching out to students rather than wait for students to come to them" (Rowh, 2018, p. 31). The CSA, funded with FF-ABE Dollars, delivered intense academic instruction to students to prepare them for their college-level coursework, help them transition into the college culture, and provide them with career information geared to programs available at the college. The CSA was a rolling program offered every 8 to 10 weeks throughout the year, with registrations occurring at similar intervals a week or two prior to the start date for each CSA session (personal communication, February 18, 2021).

Until the March 2020 move to remote-only services, students who needed to take the TSIA test did so at the college's testing center. To avoid possible student exit, testing center staff attempted to administer the TABE 11/12 tests as soon as possible after a student received a TSIA ABE Diagnostic score of a level 4 or below, which was typically either on the same day as or within a few days of the TSIA administration. When TSIA and TABE 11/12 testing resumed remotely in June 2020, TSIA testing was scheduled individually by the student. New students had to schedule a remote orientation with an advisor, and the advisor then worked with the student to book a remote TABE 11/12 testing appointment. Students received no instruction during their enrollment process into the college success academy (personal communication, February 2021). Closely timed administration of the test instruments made possible this investigation into the relationships of the tests through correlational and regression analyses on scores and placements from the TSIA-M, TSIA-M ABE Diagnostic, and TABE-M.

Participant Data Set

The study site provided me via secure email four Excel spreadsheets with the following variables and data:

- The first file included TABE 11/12 scores and test administration dates for students who enrolled in the FF–ABE’s College Success Academy between January 2019 and January 11, 2021. The variables were the TABE 11/12 form, TABE 11/12 subject area, TABE 11/12 score, and TABE 11/12 assessment date.
- The second file listed TSIA and TSIA ABE Diagnostic scores and placement levels from all students who took at least one TSIA or TSIA ABE Diagnostic test in any subject area between January 2019 and January 11, 2021. Variables consisted of a unique record identifier, the TSIA subject area, the TSIA score, the TSIA ABE Diagnostic subject area, the TSIA ABE Diagnostic placement level, and the TSIA assessment date.
- A third file included student demographic information for students. Variables included a unique record identifier, student age at enrollment, student gender, student race (Alaskan or Native American, Asian, Black or African American,

Hawaiian or Pacific Islander, Mix of two or more races, or White), and student ethnicity (Hispanic or non-Hispanic).

- A file with field notes provided information on all variables from the three Excel spreadsheets.

Variables

For correlational analyses, variables are not considered dependent or independent. However, because I originally planned to conduct logistic regression to answers research questions 3 and 4, I consistently treated the TSIA-related variables as independent and plotted them on the x-axis. Conversely, I treated the TABE 11/12-related variables as dependent and plotted them on the y-axis.

- The TSIA-M score was a continuous, scale interval variable ranging from 310 to 390. However, because students only students who scored below a 336 on the TSIA-M were administered the TSIA-M ABE Diagnostic, the range of TSIA-M scores was restricted to between 310 and 335.
- The TSIA-M ABE Diagnostic level was an ordinal variable ranging from a 1 to a 6.
- The TABE-M score was a continuous, scale interval variable ranging from 300 to 900.
- The TABE-M level was an ordinal variable ranging from a 1 to a 6.
- Gender was a dichotomous categorical variable indicating female or male.
- Ethnicity was a dichotomous categorical variable indicating Hispanic or non-Hispanic.

- Race was a categorical variable indicating Black or African American, White, or mix of other races.

Instrumentation

Instruments used for this study included two standardized academic skills assessments, the mathematics version of the TSIA and the mathematics sections of the TABE 11/12 version 11/12. I focus the following discussion on the technical aspects of both tests' reliability and validity on the publishers' program or technical manuals they provided me.

Texas Success Initiative Assessment (TSIA)

The TSIA was an adaptation of the College Board's ACCUPLACER that aligned with the Texas College and Career Readiness standards (College Board, 2014). TSIA testing commenced in fall 2013 and was replaced by an updated version, the TSIA 2.0, on January 11, 2021 (19 Tex. Admin. Code §4.56). The TSIA system consisted of tests in three academic areas: reading comprehension, mathematics, and writing. According to THECB (2014) the TSIA:

> Provides a single, statewide college-readiness threshold and a diagnostic profile by subject area for students not meeting that threshold. Students not meeting the threshold are further classified as demonstrating knowledge and skill levels at developmental education (DE) or adult basic education (ABE) levels. DE, guided by the National Reporting System for Adult Education Educational Functioning Level Descriptors . . . is now defined in Texas as knowledge and skill levels at ninth n twelfth grade (levels 5–6). ABE is defined as knowledge and skill levels at first through eighth grade (levels 1–4). (p. 33)

The College Board developed the TSIA based on item response theory, and TSIA college readiness tests for the three content areas were administered as untimed computer adaptive tests (College Board, 2014). Examinees' responses to 20–24 test items determined a scale score of 310 to 390. For each subject area, the TSIA immediately launched into one of two diagnostic assessments for students who initially tested below college level, which for mathematics was a scale score below 350. For example, for mathematics, a student's TSIA score between 336 and 349 triggered the immediate administration of the mathematics TSIA DE diagnostic module, whereas a student TSIA score below 336 triggered the immediate administration of the mathematics TSIA ABE Diagnostic and placement module. After completion of either the DE diagnostic or ABE Diagnostic and placement module, students received a score report that indicated their areas of strengths and weaknesses (Morales-Vale & Montognese, 2017).

Because the TSIA determined the student's placement into DE, the DE diagnostic module provided diagnostic information only. In contrast, the TSIA-M ABE Diagnostic assessment served both as a diagnostic and a placement tool. The TSIA-M ABE Diagnostic modules consisted of 60 questions separated into the following four strands: (a) number sense; (b) patterns, functions, and algebra; (c) statistics and probability; and (d) geometry and measurement. Students received a placement score on the ABE Diagnostics that ranged between 1 and 6. These placement scores, in turn, aligned with NRS levels (THECB, 2014). NRS levels of a 5 or 6 qualified the student for placement into DE, which up until fall 2021 could be a non-course-based option, a stand-alone class, or a corequisite DE class that was taken concurrently with the corresponding English or mathematics gateway class. THECB recommended placements for serving students who placed at ABE levels 1 through 4 (Morales-Vale & Montognese, 2017, p. 21). Figure 2 in Chapter 2 illustrates the path through the mathematics TSIA and depicts the required and allowable placements based on the student's college ready, DE, or ABE levels.

Reliability and Validity of the TSIA

The College Board (2014) released a program manual for the TSIA that outlines the test development process and assured ongoing analysis of student test scores. To determine cut scores, the College Board convened adult education, developmental education, and college-level instructors to participate in a modified bookmark method that established cut scores for college-ready, developmental education, and adult education NRS levels. The College Board also reported that it would be continuously performing functioning differential item analysis to ensure that no one group of students had an advantage or disadvantage over another; items flagged would be analyzed and revised or reviewed if they may be biased (College Board, 2014). Although the College Board states that rigorous testing has been performed to ensure the validity and reliability of the TSIA, there has been no public release of a technical report (S. Morales-Vales, personal communication, April 25, 2021).

THECB required that the College Board conduct research to explore its accuracy at placement students into college-level coursework. Cui and Bay (2017) concluded that the placements resulting from TSIA college readiness cut scores sufficiently predicted success in the student's mathematics, reading-intensive, and English gateway courses. Using logistical regression modeling, the authors calculated that students with minimum cut scores of 350 for mathematics had a 67% probability of earning a grade of a C– or above in their first mathematics course. For a minimum cut score of 351 in reading, students had a 64% chance of earning a C– or above in their first reading-intensive course. TSIA writing scores are combined with an essay test, so students with multiple-choice writing test scores of 350 and essay scores of 5 had a 74% probability of earning a C– or above in their first English course. A regression model was not found that fit the combination of a TSIA writing score of 350 and an essay score of 4, which prompted THECB to adjust the college readiness cut scores for writing.

Cui and Bay (2017) also analyzed the frequencies of correct placement in college-level coursework along with error rates resulting from underplacement and overplacement of students based on their TSIA scores. Correctly placed students were those who met minimum cut scores, were placed in college-level gateway courses, and passed them with a C– or above. Overplaced students were those who met minimum cut scores, were placed in college-level gateway courses, but either withdrew or earned a grade of a D or an F. Underplaced students were students who did not meet minimum cut scores but who may have been exempted through other assessment scores, veteran status, or other reasons and were allowed to enroll in the college-level gateway course and did pass the course with a grade of C– or above. For mathematics, the correct, overplacement, and underplacement percentages were 63%, 23%, and 15%, respectively. For reading, the correct, overplacement, and underplacement percentages were 69%, 21%, and 11%, respectively. For students with a multiple-choice writing score of 350 or above with an essay score of 5 or above, 75% were correctly placed, 18% overplaced, and 7% underplaced. Overall, the results of Cui and Bay's (2017) study of the predictive validity of the TSIA on student gateway course success are similar to the

ACCUPLACER meta-analysis predictive validity study conducted by Mattern and Packman (2009).

Tests of Adult Basic Education (TABE 11/12)

The TABE 11/12 system consists of tests in three academic areas: reading comprehension, language, and mathematics. The purpose of the TABE 11/12 is to measure the academic skills of adults. Initially developed in 1967, there have been seven revisions to the TABE 11/12 (Johnson, 2019). In 2015, Data Recognition Corporation (DRC) purchased rights to the TABE 11/12, and in October 2017, DRC submitted the TABE 11/12 to the U.S. Department of Education for approval. The NRS authorized the TABE 11/12 for use in FF-ABE programs beginning September 2017. In Texas, the state ABE office required all programs to use the revised TABE 11/12 beginning with the July 2019 program year (Johnson, 2019).

The TABE 11/12 system comes in two forms, the TABE 11 and the TABE 12, a locator test, and five levels of the test form that are progressively more difficult. Students must first be assessed with the TABE 11/12 locator to determine their approximate test level. The five levels of the test correspond to grade equivalencies (GE) and NRS level placements as follows:

- Literacy (L), NRS levels 1–2, GE 0–1.9
- Elementary (E), NRS levels 1–3; GE 2–3.9
- Medium (M), NRS levels 2–4; GE 4–5.9
- Difficult (D), NRS levels 3–5; GE 6–8.9
- Advanced (A), NRS levels 4–6; GE 9–12.9

Within the TABE 11/12 system, there are a total of 36 subtests: a locator test for each test form in all three subject areas (totaling six tests), with five test levels for each of the two forms in all three subject areas (totaling 30 tests). All tests are multiple-choice and based on item response theory. The print-based version is timed, whereas the computer-based version is adaptive.

The grade-equivalency concordance of the TABE 11/12 was based on research with the California Achievement Test (Rogers, 1988, as cited in Monsaas, 2007). When interpreting GE in Table 10, the integer is the school grade level, and the decimal is the month of the nine-month school year. Therefore, a 7.4 GE indicates an academic level similar to that of the fourth month of the seventh grade (CTB/McGraw-Hill, 2004). However, the use of GE may be somewhat confusing because it must be interpreted within the difficulty level of the test. For example, a student who takes a level D test that is designed for GE 6–8.9 and scores a 665, which corresponds to a GE of 10.9, that student is not performing at the level of a student in the ninth month of eighth grade. Rather, the interpretation is that the student is performing as would a student in the ninth month of eighth-grade who is taking a test on middle school level academic skills (Truttschel, 2010).

Reliability and Validity of the TABE 11/12

Ensuring that a test measures what it is supposed to measure, which is the text construct, begins with thorough knowledge of the content the test will measures. To develop the TABE 11/12, DRC aligned the test’s content and rigor to the federal adult education content standards, which were developed as a subset of the Common Core State Standards for college readiness (Pimentel, 2013). The Common Core State

Table 10

TABE 11/12 Cut Scores with Corresponding NRS Placements and Grade Equivalencies

TABE 11/12 Cut Scores	NRS Level	NRS Level Name	GE [a]
300–448	1	Beginning ABE Literacy	0. 0–1. 9
449–495	2	Beginning Basic Education	2. 0–3. 9
496–536	3	Low Intermediate Basic Education	4. 0–5. 9
537–595	4	High Intermediate Basic Education	6. 0–8. 9
596–656	5	Low Adult Secondary Education	9. 0–10. 9
657–800	6	High Adult Secondary Education	11. 0–12.9

Note. Adapted from Test Benchmarks for NRS Educational Functioning Levels (NRS, 2019). [a] GE is the grade level equivalency of the NRS Level.

Standards (CCSS) evolved from a coalition of policymakers from 48 states who collaboratively developed college and career readiness standards for high school exit and have adopted as the basis of their K-12 content standards (Common Core States Initiative, 2018). Shortly after the release of the CCSS, the U.S. Department of Education created the College and Career Readiness Standards for Adult Education (CCRS-AE) to provide FF-ABE practitioners a set of English Language Arts and mathematics benchmarks that ranged from low literacy to college readiness levels. According to Pimentel, 2013, "representatives from adult education, community colleges, career and technical training, and the military” (p. 7) deliberated for several months to agree on the selection of a subset of standards from the CCSS that were most relevant for FF-ABE students.

Because the TABE 11/12 was created to measure *College and Career Readiness Standards for Adult Education* (DRC, 2017), it was crucial for the TABE to determine accurate cut scores to classify students into one of the National Reporting System of Adult Education's (NRS) six educational function levels (NRS levels). To do this, DRC convened 600 adult education practitioners from community college, K-12, community, and corrections settings for a bookmark standards-setting procedure. Prior to the standards-setting event, DRC placed the TABE 11/12 test items in order of difficult using a test linkage between the TABE 11/12 and TerraNova, a K-12 assessment system. During the bookmark event, for NRS levels 2 through 6, participants individually flagged the test items they believed at least 67% of adult learners would not complete correctly. Participants then met to discuss their responses in small groups, after which they repeated the bookmark process. At this point, the most common test item that determined entry into NRS levels 2 through 6 were provided to two smaller groups of content experts who were tasked with setting the cut off scores. To assist with this, DRC also gave these two groups impact statements that estimated the percentage of students who would place into each level for each cut score. Once this process was complete, DRC assembled a review committee to recommend any adjustments. The final process involved minor changes to ensure that each NRS level had a sufficient range of scale scores (DRC, 2017).

Because the TABE 11/12 was designed as a criterion-referenced assessment, Data Recognition Corporation (DRC) conducted extensive norming studies to assure that test performances across various groups were comparable (DRC, 2019). The TABE 11/12's national norming sample included data from more than 34,000 examinees from 400 institutions in 46 states, with students in ABE, adult secondary education, ESL programs, alternative high schools, juvenile and adult correction facilities, and vocational/technical programs. Also, prior to the field testing, DRC worked with item writers to reduce gender, race, or ethnicity bias in test items. After the field testing, DRC conducted a differential item functioning analysis using the Mantel-Haenzel to ensure that no test items were biased against a test taker based on race or gender. This analysis resulted in the removal of 27 items that favored either gender or race across the 30 tests (DRC, 2017).

The TABE 11/12's internal consistency reliability for scores was determined using Cronbach's alpha and item response theory (IRT)-based reliability estimates. The results for both types of analyses were similar. With the exception of three mathematics forms and one language form, the 30 TABE 11/12 subtests displayed Cronbach alpha and IRT-based reliability estimates of .80 or above, with the values ranging from .75 to .79 for the remaining four subtests. DRC conducted several additional procedures to establish the validity of the TABE 11/12. For example, in 2016, the publisher conducted concurrent validity and score linking research using a sample of 3,000 students who were administered both the older TABE 9/10 along with the TABE 11 or TABE 12 during field testing. After correcting for attenuation for test measurement error, correlations between TABE 9/ 10 and TABE 11/12 subtests ranged between .57 and .85, with only five subtest correlations under .70. According to DRC, these strong correlations supported the construct validity of the TABE 11/12 (DRC, 2017).

DRC conducted concurrent validity and score linking research between the advanced levels of the TABE 11/12 and a standardized college readiness test, the Tests Assessing Secondary Completion (TASC). The publisher used sample sizes of over 250 participants for each subtest, which participants took concurrently. The uncorrected correlation between the TABE 11/12 advanced mathematics and the TASC mathematics was .62, and the correlation between the TABE 11/12 advanced reading and the TASC reading was .64. Correlations corrected for sample size and range restriction were .72 between the mathematics subtests and .69 between the reading subtests. DRC concluded that these correlation figures indicate that the two tests measure similar constructs and that the correlations build support for the TABE 11/12's validity (DRC, 2017).

One of the major performance measures for FF-ABE programs is the percentage of students in the program who make a gain in an NRS level from a baseline and a progress test. Therefore, as part of validity and reliability evaluation, DRC evaluated the placement accuracy between student NRS placements levels on the TABE 11/12 forms 11 and 12 and between the TABE 11/12 computer-based versus print-based test forms using three techniques: (a) Calculation and graphing of the IRT conditional standard error of measure (CSEM) within each test level, (b) standard error of calculation and interpretation of Cohen's kappa, and (c) consistency of classification. The plotted relationships from the TABE 11 and the TABE 12 student scores and CSEM were extremely similar within each test level. However, for scores that fall near the bottom or the top of the test score range, the test publisher recommends the administration of a lower or higher level TABE 11/12 test, as appropriate, to produce a more reliable measure (DRC, 2017).

DRC reported Cohen's weighted kappa, which "estimates the proportion of consistent classifications between two forms by removing the proportion of consistent classifications that would be expected from chance alone" (Cohen, 1988, as cited in

DRC, 2017, p. 85). The weighted kappa value for the fifteen TABE 11/12 subtests ranged from .51 to .73, indicating "moderate to substantial agreement (p. 85)" in NRS level placement accuracy.

The third technique to establish reliability and validity of the TABE 11/12's NRS placements was a consistency of classification analysis in which students were administered both forms of the tests and percentages of placement agreements and differences were calculated. The consistency of classification analysis for NRS placement levels made by the TABE 11 and the TABE 12 averaged 66.2% across the five levels of math tests. According to DRC, reliability and validity for placements from two different instruments doubles the amount of measurement error. In addition, the number of levels in a scale, the number of items at each level, and the number of examinees who score near the level cut-off points will also affect classification accuracy rates. For this reason, placement accuracy of at least 50% is acceptable, with higher levels of agreement preferred (DRC, 2017).

Data Analysis

The section that follows describes the types of data analysis that I performed to answer my research questions. I used Spearman's r_s for Research Questions 1 and 2, Fisher's r-to-z transformation, and Spearman's r_s standardized Fisher's *z*-value confidence intervals for Research Questions 3 and 4, and analysis of consistency of classification for Research Question 2.

Spearman's r_s

To test the null hypotheses associated with Research Questions 1 and 2, I calculate a Spearman's r_s correlation coefficient. Research Question 1 asked if there was a relationship between student scores on closely times TSIA-M and the TABE-M assessments. The null hypothesis stated that there is no statistically significant relationship between the scores on the two assessments. Variables I used for Research Question were the TSIA-M scores and the TABE 11/12-scores. The TSIA-M scores were restricted in range to only 310–335 out of a high score of 390. Furthermore, many students had mathematics skills that were lower than the TSIA-M was designed to measure, so the distribution was extremely negatively skewed, with many students scoring a 310. Although I attempted to transform the data to a more normal distribution, this was unsuccessful. Therefore, even though both variables were interval in nature, I chose to use a Spearman's r_s as a nonparametric test of correlation.

I also chose to calculate Spearman's r_s correlation coefficient to answer Research Question 2. Research Question 2 asked if there was a relationship between student NRS placements on the TSIA-M ABE Diagnostic and the TABE-M. T The null hypothesis stated that there is no statistically significant relationship between student placements on the two assessments. Both the TSIA-M ABE Diagnostic and the TABE-M align to the NRS ABE education functioning levels, which uses an ordinal scale from 1 to 6. The presence of one or more ordinal variables in a correlation indicates the use of a nonparametric test such as the Spearman's $r_{s.}$

Fisher's *r-to-z* Transformation

Research Questions 3 asks if controlling for age, gender, race, or ethnicity changes the relationships between student scores on closely administered TSIA-M and TABE-M assessments. The associated null hypothesis was that there are no statistically significant differences in the relationships between student scores on closely administered mathematics TSIA-M and TABE-M assessments after controlling for age, gender, race, or ethnicity. My interest was to determine if there were differences between the relationships on student performance based on membership in a particular student group. Therefore, I needed to determine the correlations between the two test scores for each subgroup and then find a procedure that could see if there were significant differences between those relationships.

For dichotomous variables, the Fisher's *r-to-z* transformation calculates if there is a significant difference between two correlations. I was able to use this to compare the dichotomous variables male and female and Hispanic and non-Hispanic. Because my categorical variables of age group and race had three values, a direct Fisher's *r-to-z* transformation was not possible. Instead, I was able to convert the correlations into *z*-values with confidence intervals that could be examined using an SPSS script provided by Gignac (n.d., 2019b). A *z*-value can fall anywhere in the confidence interval range, so if two variables have overlapping confidence intervals, they could have the same values. If the variables have the same values, then they are not significantly different. If the results of the Fisher's *r-to-z* transformations are significant, then the correlations between the test scores from one group are significantly different than other groups. With assessment, it is important to test for group differences so that they become known. If the tester is aware of a known group difference, it may change the way the score should be interpreted or used.

Research Questions 4 asks if controlling for age, gender, race, or ethnicity changes the relationships between student placements on closely administered TSIA-M and TABE-M assessments. The associated null hypothesis was that there are no statistically significant differences in the relationships between student placements on closely administered mathematics TSIA-M and TABE-M assessments after controlling for age, gender, race, or ethnicity. My interest in exploring this question was the same as for Research Question 3. I used the same procedure, which was to calculate Spearman's r_s coefficients. For dichotomous variables, I calculated if there was a significant difference in their correlations by using Fisher's *r to z* transformation. For categorical variables with 3 values, I transformed the correlations to Fisher *z*-values and calculated the confidence intervals for each z-value to determine if there were differences between the relationships on student performance based on membership in a particular student group. I then plotted the confidence intervals to analyze relationships between the confidence intervals to determine if there were any significant differences in correlations across demographic groups.

Consistency of Classification Analysis

For Research Question 2, I performed analysis to determine the accuracy of the TSIA-M ABE Diagnostics as compared to the TSIA-M. I did so because more important than a strong correlation between placements is their accuracy at classifying students in the same NRS levels. Ultimately, placement is the intended use of the TSIA-M ABE Diagnostic. From my initial process of inquiry, I have been curious to know if a pattern existed of more or less consistent overplacement or underplacement of the TSIA-M ABE Diagnostic as compared with the TABE-M's placements. Through this analysis, I expect to gain knowledge that may be practical in nature and useful to practitioners who work with students who are in the process of postsecondary transition or who need placement into instructional options that are best suited to their needs.

Chapter Summary

In this chapter, I began by describing my research design, which is correlational in nature and is similar to research that looks for evidence of convergent validity by comparing an unproven measurement to a well-established one in an effort to indirectly verify the construct validity of the unproven instrument. I also presented my research questions and null hypotheses. I provided information on my IRB process and gave a description of the local setting that included the procedures that were used to test students with the TSIA-M, TSIA ABE Diagnostic, and TABE-M. I delineated the data and variables I received from the research site and defined each variable that I used in the study.

Because I assumed that the TABE 11/12 was the criterion against which the TSIA-M and TSIA-M ABE Diagnostic would be compared, I included extensive technical information for the TABE 11/12. I also explained in detail the process involved with the TSIA administration and placements that result based on TSIA ABE Diagnostic results. After listing my research questions and associated null hypotheses, I outlined the statistical testing I went through to test the null hypotheses, which included calculating Spearman's r_s, Fisher z-values, Fisher's *r-to-*z transformation, and consistency of classification analysis. In the next chapter, I will again explain each analysis and its results.

IV: RESULTS

The purpose of this investigation was to compare student performance on closely administered TSIA-M, TSIA ABE Diagnostic, and TABE-M assessments by analyzing the relationships between student scores and NRS placement levels on these instruments. In this chapter, I discuss the results of the data analysis. I first describe the data I received and the procedures I used to prepare the data set. Next, I present descriptive data for all variables. I then answer each research question by choosing an appropriate statistical test whose results indicated the rejection or acceptance of the corresponding null hypotheses for each research question. I conclude this chapter with a summary that includes information on whether I rejected or failed to reject the null hypothesis based on the data analysis.

Data Preparation

I received four Excel files from the study site. The file with scores for the Test of Adult Basic Education 11/12 (TABE 11/12) scores included TABE 11.12 scores and test administration dates for students who enrolled in the adult education College Success Academy between January 2018 and January 2021. The file with Texas Success Initiative Assessment (TSIA) scores listed TSIA mathematics (TSIA-M) scores and TSIA-M adult basic education (TSIA-M ABE) diagnostic placement levels from all students who took at least one TSIA-M and one TSIA-M diagnostic test in any subject area between January 2018 and January 2021. In the student information file, I received demographic information for students, including age at enrollment, gender, race, and ethnicity. A file with field notes provided information on all variables from the three data files.

To create my final data set, I first used the information provided on the TABE-M scores to filter out only students who were administered the mathematics TABE form 11 or 12, which yielded 176 records with 164 mathematics scores, 69 reading scores, and 49 language scores. Although my initial plan included an analysis of student performance on the assessments for all three subject areas, the sample sizes for reading and language were insufficient. Therefore, I limited my investigation to a comparison of student performance on the mathematics sections of the TSIA-M, the TSIA-M ABE Diagnostic, and the TABE-M.

I created the final data set by matching the TABE-M scores with the student demographics, TSIA-M scores, and TSIA-M ABE Diagnostic levels. I removed three students who were missing TSIA-M ABE Diagnostic level data and nine students who took the TSIA-M and TABE-M assessments outside of the 2–week time frame I used as a definition of closely-timed test administrations. The final data set consisted of 152 records. I manually added the TABE-M NRS levels to the data using the Test Benchmarks for NRS Educational Functioning Levels (NRS, 2019). Finally, before analyzing the data set's descriptive statistics, I recoded student demographic information into variables with nominal or ordinal values.

Descriptive Statistics

Compared to the study site, the demographics of students who were administered the TSIA-M ABE Diagnostic was similar except for an overrepresentation of female students and of African American students by approximately 10%. Missing data on the demographic characteristics included 20 students who did not report race and one student who did not report ethnicity (see Table 11). The sample size, mean, standard deviation, and range for each of the assessments used for this investigation appear in Table 12, with both the observed and possible range of scores and NRS placement levels listed for each assessment. Two TSIA-M scores and five TSIA-M ABE Diagnostic placement levels were missing from the data. Although the standard deviation appears smaller for the TSIA-M, the mean score is lower and the range of possible scores on the test is much narrower than the range for the TABE 11/12's scores. Finally, Table 13 displays a comparison of the percentages of students who placed into NRS levels via the TSIA-M ABE Diagnostic and the TABE-M. Comparing the mean NRS placement levels and the percentages placed into each NRS level by the two assessments indicated the TSIA-M Diagnostic placed students into higher NRS levels than did the TABE-M.

Table 11

Student Demographics Descriptive Statistics

Student Characteristics	*n*	%
Age		
18–21	98	64%
22–29	25	16%
30 and Over	29	19%
Gender		
Female	108	71%
Male	44	29%
Race		
African American	41	31%
White	79	60%
Other	12	9%
Ethnicity		
Non-Hispanic	93	62%
Hispanic	58	38%

Table 12

Sample Sizes, Means, Standard Deviations and Ranges for Student Assessment Scores and Levels

Variable	*n*	*M*	*SD*	Observed Range	Test Range
TSIA-M Scores	150	318.15	8.25	310–335	310–390
TABE-M Scores	152	500.88	28.10	401–572	300–800
TSIA-M ABE Diagnostic Levels	147	3.84	.61	1–5	1–6
TABE-M Levels	152	2.71	.75	2–5	1–6

Note. The Observed Range is the range of scores from study participants. The Test Range is the possible range of scores or levels on the tests.

Table 13

TSIA-M ABE Diagnostic and TABE-M Student NRS Level Placements

Assessment Instrument	NRS Levels									
	Level 1		Level 2		Level 3		Level 4		Level 5	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
TSIA-M ABE Diagnostic	0	0%	6	4.1%	23	21.5%	107	72.8%	11	7.5
TABE-M	4	2.6%	56	36.8%	75	49.3%	14	9.2%	3	2.0%

Results of Null Hypothesis Testing

In the following section, I first list each research question and corresponding null hypothesis. Next, I outline the statistical procedures I used and how I test the procedure's assumptions. Finally, I report if I rejected or failed to reject each null hypothesis based on the results of the statistical analyses performed.

Research Question and Null Hypothesis 1

- What is the relationship between student scores on closely administered mathematics TSIA-M and TABE-M assessments?
- H0: $r_s = 0$; there is no statistically significant relationship between student scores on closely administered mathematics TSIA-M and TABE-M assessments.

To answer this question, I first explored my data. First, the data met the assumption of related pairs. Next, I used *SPSS* to explore the distributions of TSIA-M and TABE-M scores, which the histograms in Figures 3 and 4 illustrate. Per *SPSS* calculation of skewness and kurtosis, I determined that the TABE mathematics scores were normally distributed, but the TSI mathematics scores were extremely negatively skewed. Visual inspection also indicated one probable outlier in the TSIA mathematics scores and three probable outliers in the TABE mathematics scores.

The asymmetrical distribution of the TSI mathematics scores required me to treat the scores as non–interval. Therefore, I used *SPSS* to run a non–parametric Spearman's rank-order correlation between the TSIA-M and TABE-M mathematics scores. Spearman's rank-order correlation coefficient value r_s can be interpreted to test the strength of association between ordinal or interval variables. Spearman's correlation is recommended for interval variables when there are non-normal data distributions in one

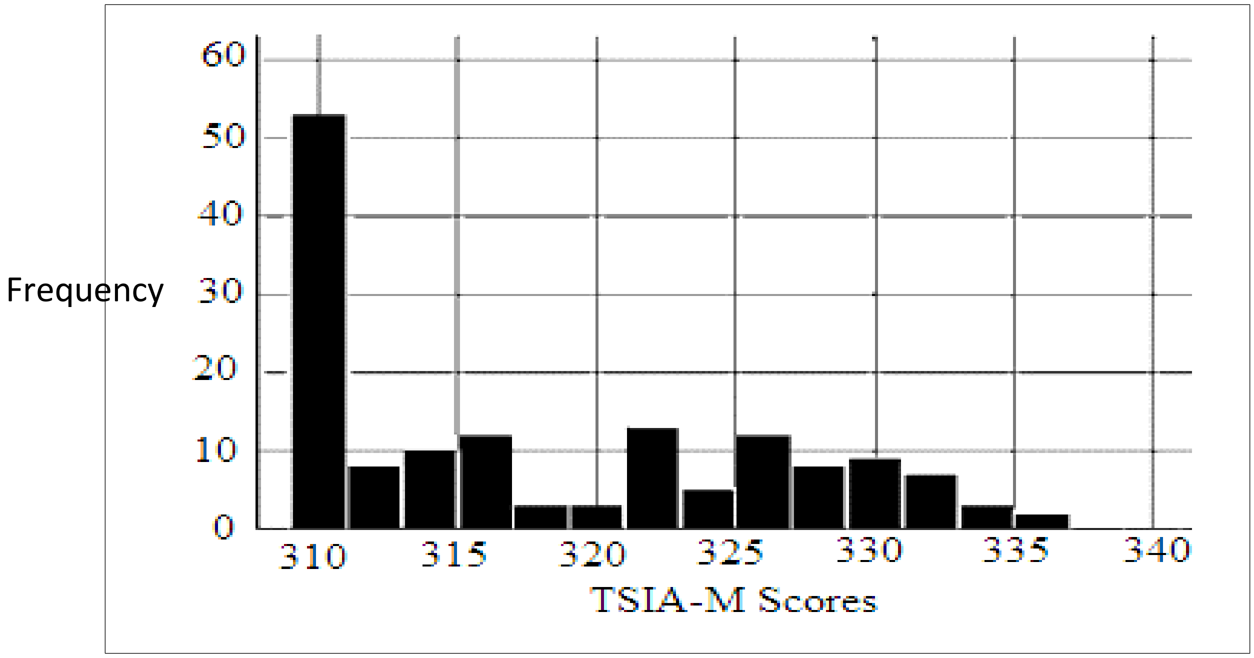

Figure 3. Frequency Distribution of TSIA-M Scores

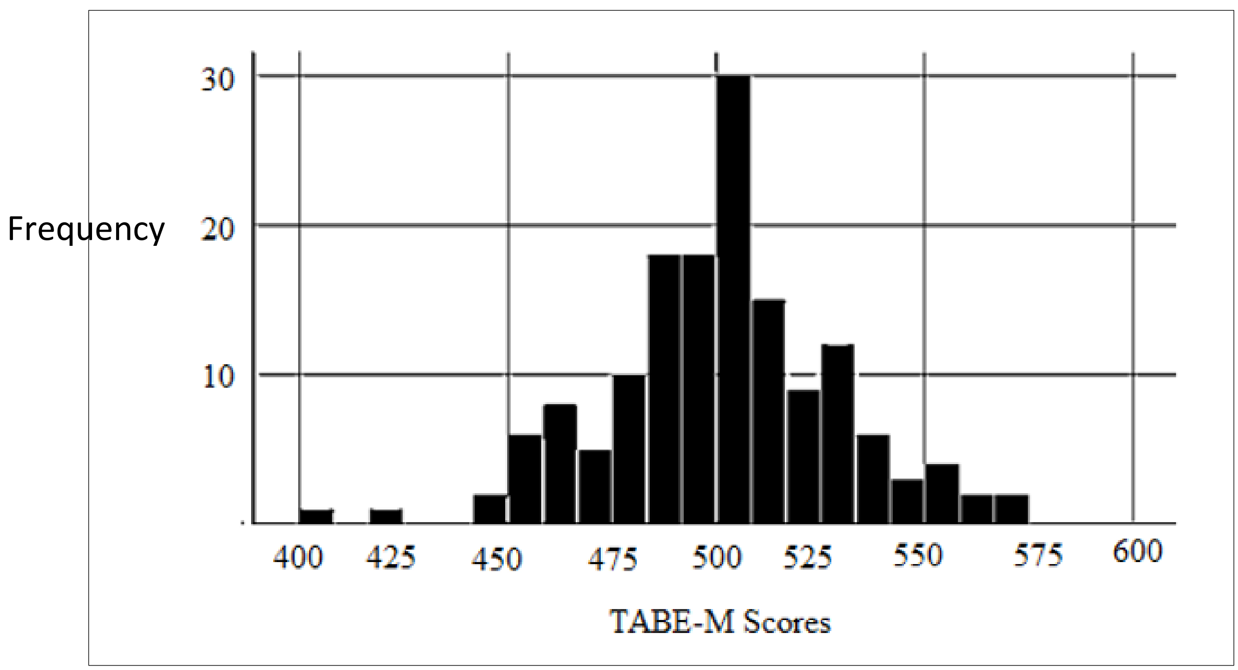

Figure 4. Distribution of TABE-M Scores

or both variables (Gignac, 2019a, Chapter 5). Rather than using the actual numerical values, Spearman's correlation ranks the values of the variables and then correlates the ranked values. Both variables met the assumptions for Spearman's r_s, which permits either two ordinal variables or an ordinal and continuous variable to be correlated if the relationship is monotonic, which I verified by visual inspection of a scatterplot between the variables as shown in Figure 5. Because non–parametric correlations are less affected by outliers (Gardner & Altman, 2000), I chose to leave the probable four outlying cases in the data set.

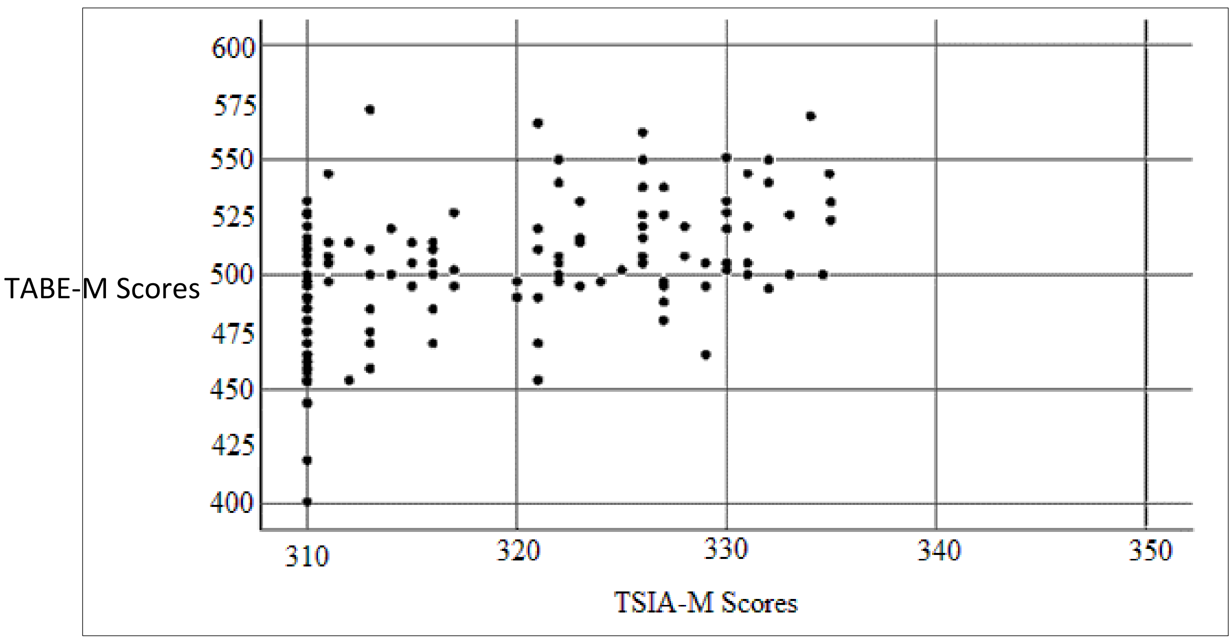

Figure 5. *Scatterplot of TSIA-M Scores and TABE-M Scores*

Results of the Spearman's correlation indicated a statistically significant positive correlation between TSIA-M and TABE-M scores; r_s (150) = .542, $p < .001$. According to Cohen's (1988) recommended interpretation of the strength of association for Spearman's r_s, the correlation is considered strong. Therefore, I rejected the null hypothesis and concluded that a strong, statistically significant relationship existed between student scores on the TSIA-M and TABE-M.

Research Question and Null Hypothesis 2

1. What is the relationship between NRS student placements on closely administered TSIA-M ABE Diagnostic and TABE-M assessments?
2. H0: $r_s = 0$; there is no statistically significant relationship between NRS student placements from closely administered TSIA-M ABE Diagnostic and TABE-M assessments.

This question required a correlation between the NRS level placements from the TSIA-M ABE Diagnostic and TABE-M assessments. The variables were both ordinal, so I again chose Spearman's correlation, a test often used to discover the strength of association between ordinal variables (Gardner & Altman, 2000). The scatterplot in Figure 6 illustrates the monotonic relationship between the TSIA-M ABE Diagnostic and TABE-M scores.

Results of the Spearman's correlation indicated a statistically significant positive correlation between the TSIA-M ABE Diagnostic and TABE-M NRS level placements; $r_s(147) = .337$, $p = .01$. The strength of the correlation for the Spearman's test is moderate. Therefore, I rejected the null hypothesis and concluded that a positive, statistically significant relationship of moderate strength existed between students' TSIA-ABE Diagnostic and TABE-M student placements.

To further explore the level of NRS level agreement between the TSIA-MD and the TABE-M, I engaged in a consistency of classification analysis. I created a new variable, LEVEL_DIFF, by subtracting students' TABE-M levels from their TSIA-

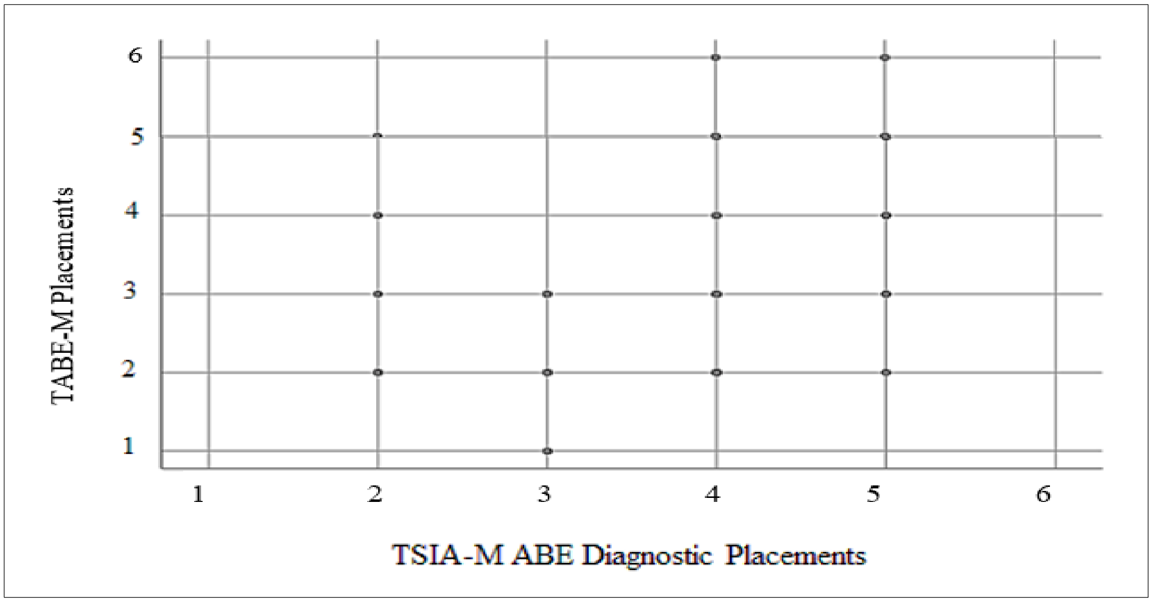

Figure 6. *Scatterplot of TSIA-M ABE Diagnostic and TABE-M Placements*

M ABE Diagnostic levels. The average difference between placement levels was –1.0, and the standard deviation was 1.013. Only 15% of the students were placed into the same level by both assessments. The TSIA-M ABE Diagnostic placed over half of the students one level higher than the TABE-M and two levels higher for over almost one-third of the students (see Table 14). Figure 7 illustrates the frequency distribution of these placement differences, which is biased for student placement at a higher level on the TSIA-M ABE Diagnostic than on the TABE-M.

I also calculated the consistency of classification based on the overplacement, exact placement, and underplacement for each TSIA-M ABE Diagnostic level placement, as illustrated in Table 15. For levels 3 through 5, there were no underplaced students, whereas for level 2, there were no overplaced students. Approximately 70% of all students placed into NRS level 4 on the TSIA-M ABE, which had one of the higher exact consistency of classification rates of almost about 30%.

Table 14

Frequency Table: Differences in TSIA-M ABE Diagnostic and TABE-M Placements

NRS Level Difference [a]	NRS Level Difference Frequency	NRS Level Difference Percent
3	1	0.7
2	44	29.9
1	77	52.4
0	22	15.0
–1	2	1.4
–2	1	0.7
Total	147	100.0

Note. [a] NRS level difference is the difference between the TSIA-M ABE Diagnostic and TABE-M placement levels.

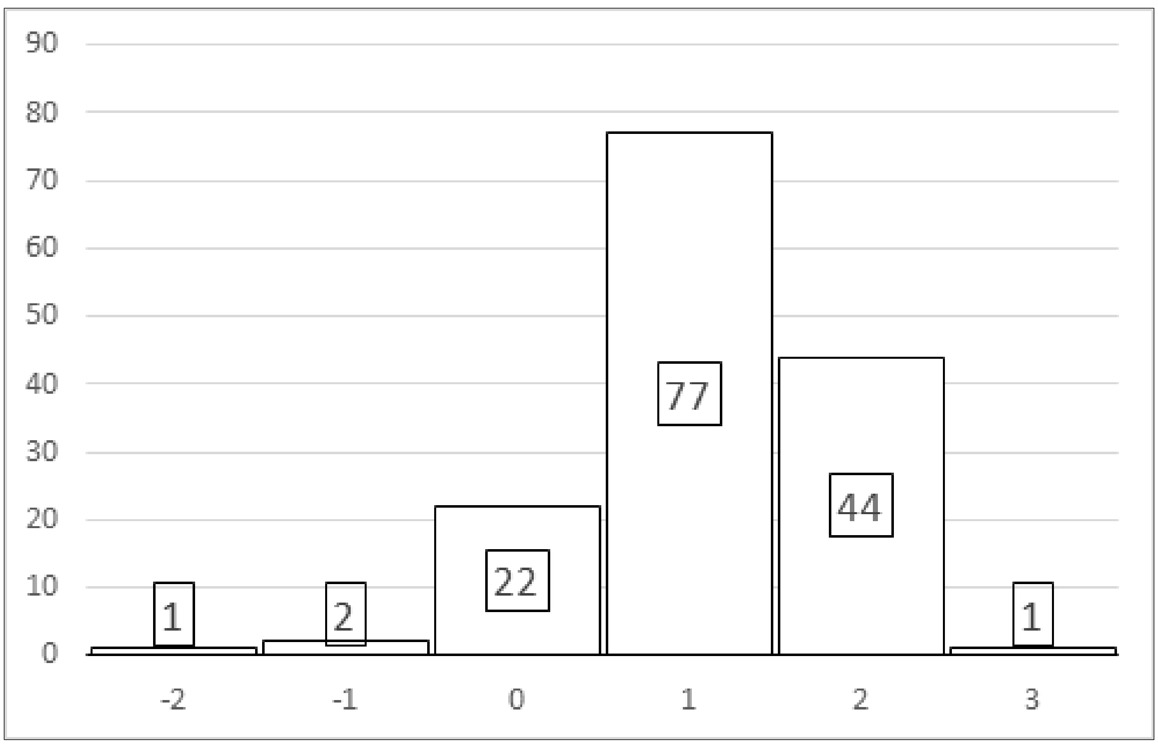

Figure 7. *Frequency Distribution: TSIA-M ABE Diagnostic Overplacement.* Differences in NRS Level Placements are the differences between the TSIA-M ABE Diagnostic and TABE-M placement levels.

Table 15

Consistency of Classification per TSIA-M ABE Diagnostic Placement Level

TSIA-M ABE Diagnostic Placement	Difference in NRS Placement Levels						
	n	+3	+2	+1	0	–1	–2
NRS Level 5	11	9.1	27.3	45.5	18.2	0.0	0.0
NRS Level 4	107	0.0	17.4	52.2	30.4	0.0	0.0
NRS Level 3	23	0.0	34.6	56.1	8.4	0.9	0.0
NRS Level 2	6	0.0	0.0	0.0	66.7	16.7	16.7
All NRS Levels	147	0.07	30.6	52.4	15.0	1.4	0.07

Note. The difference in NRS placement levels was calculated by subtracting the TABE-M NRS placement level from the TSIA-M ABE Diagnostic NRS placement level.

Research Question and Null Hypothesis 3

- Does controlling for age, gender, race, and ethnicity change the relationship between closely administered TSIA-M and TABE-M scores?
- H0: $r_s = 0$; there are no statistically significant differences in the relationships between student scores on closely administered TSIA-M and TABE-M assessments after controlling for age, gender, race, or ethnicity.

To answer this question, I needed to determine any difference in the relationships of TSIA-M and TABE-M scores across demographic characteristics. I first calculated the correlations between the TSIA-M and TABE-M scores for each subgroup. Although both variables are interval, the asymmetry of the TSIA-M scores indicated the use of nonparametric Spearman's correlation procedures.

After obtaining the correlations, I needed a way to compare their differences for strength and direction and determine the significance of the differences between them. For the two dichotomous variables, Gender and Ethnicity, I used Fisher's *r-to-z* transformation (Fisher's z), which is a procedure commonly used to compare Pearson's correlations from two groups to determine if the two correlations are significantly different. Fisher's z uses the following formula, in which r_1 is a Pearson's r correlation coefficient from one sample and r_2 is a Pearson's r correlation coefficient from another: $z = 1/2(\log(1 + r) - \log(1 - r)$. Although the formula above specifies Pearson's r, Spearman's r_s correlations may also be used (Pennsylvania State Eberly College of Science, 2021, Lesson 18.2).

Fisher's z-values are standardized, so for a significance level of 95%, any absolute value of z over 1.96 indicates a rejection of the null hypothesis. Table 16 displays the z-values for the correlations between female and male and not-Hispanic and Hispanic TSIA-M and TABE-M scores. Based on the results of the Fisher's *r-to-z* transformation test, there were no statistically significant differences between the TSIA-M and TABE-M score correlations between females and males or between non-Hispanic and Hispanic subgroups.

Because the variables for race and age group were not dichotomous, I needed to choose a different approach for comparing the significance of the race and age group test score correlations. One way to do this was by comparing the confidence intervals for each correlation to determine if they overlapped or if they were distinct from each other. If the confidence intervals between the demographic subgroups

Table 16

Significance of Correlations of TSIA-M and TABE-M Scores by Demographic Factor Using Fisher's r-to-z Transformation Test

Demographic Groups Compared	Fisher's z [a]
Female to Male	0.99
Non-Hispanic to Hispanic	0.20

Note. *Z*-values were calculated using an online calculator (Lowry, 2021). [a] *Z*-values greater than the absolute value of 1.96 indicate a statistically significant difference between the correlations of the two subgroups (Esri, n.d.).

overlapped, this would indicate that the correlation values could be the same, and therefore, the relationship between the demographic subgroups would not be significantly different. If there were no overlap between the confidence intervals, the correlations could not be the same value and would therefore be significantly different, resulting in a rejection of the null hypothesis.

To calculate the confidence intervals for the various demographic subgroups, I first used *SPSS* to calculate Spearman's correlations between TSIA-M and TABE-M for each demographic subgroup for race and age group. Next, I used Gignac's (n.d., 2019b) *SPSS* syntax (shown in Figure 8) to convert the Spearman's correlation coefficients to Fisher's *z*-values and to provide their 95% confidence interval values. Table 17 lists the Spearman's r_s correlation coefficients and levels of significance along with the corresponding Fisher's *z*-values and lower and upper confidence intervals. In Figure 9, I graphed the confidence intervals across the demographic characteristics. Both Table 17

```
COMPUTE Fishers_z = .50 * (LN(1 + rs) – LN(1 – rs)).
COMPUTE Fishers_z_lb = Fishers_z – (1.96 * (SQRT((1 + (rs **2/2)) / (n – 3)))).
COMPUTE Fishers_z_ub = Fishers_z + (1.96*(SQRT((1 + (rs **2/2)) / ( n – 3)))).
COMPUTE Ci95_r_lb = ((exp(2*Fishers_z_lb)) – 1) / ((exp(2 * Fishers_z_lb)) + 1).
COMPUTE Ci95_r_ub=((exp(2*Fishers_z_ub)) –1)/((exp(2 * Fishers_z_ub)) + 1).
EXECUTE.
```

Figure 8. *Gignac's SPSS Syntax Spearman's r_s to Fisher's z-Values***.** Instructions for calculating confidence intervals for Spearman's correlation coefficients are available at https://www.youtube.com/watch?v=gKHgCq5E864 (Gignac, 2019b). The SPSS syntax is available for download at https://tinyurl.com/y4e7dtvn (Gignac, n.d.).

Table 17

TSIA-M and TABE-M Score Correlations by Demographic Characteristics

Demographic Comparison	*n*	TSIA-M to TABE-M Scores Spearman's r_s	*p*	Fisher's *z* and Confidence
Age				
18–21	98	.522	<.001	.58 [.35, .66]
22–29	25	.639	.001	.76 [.30, .84]
30 and Over	29	.415	.025	.44 [.04, .69]
Race				
African American	41	.572	<.001	.63 [.28, .75]
White	79	.609	<.001	.71 [.43, .74]
Other	12	.490	.106	.54 [–.15, .84]

Note. Values in square brackets indicate the 95% confidence interval for each correlation.

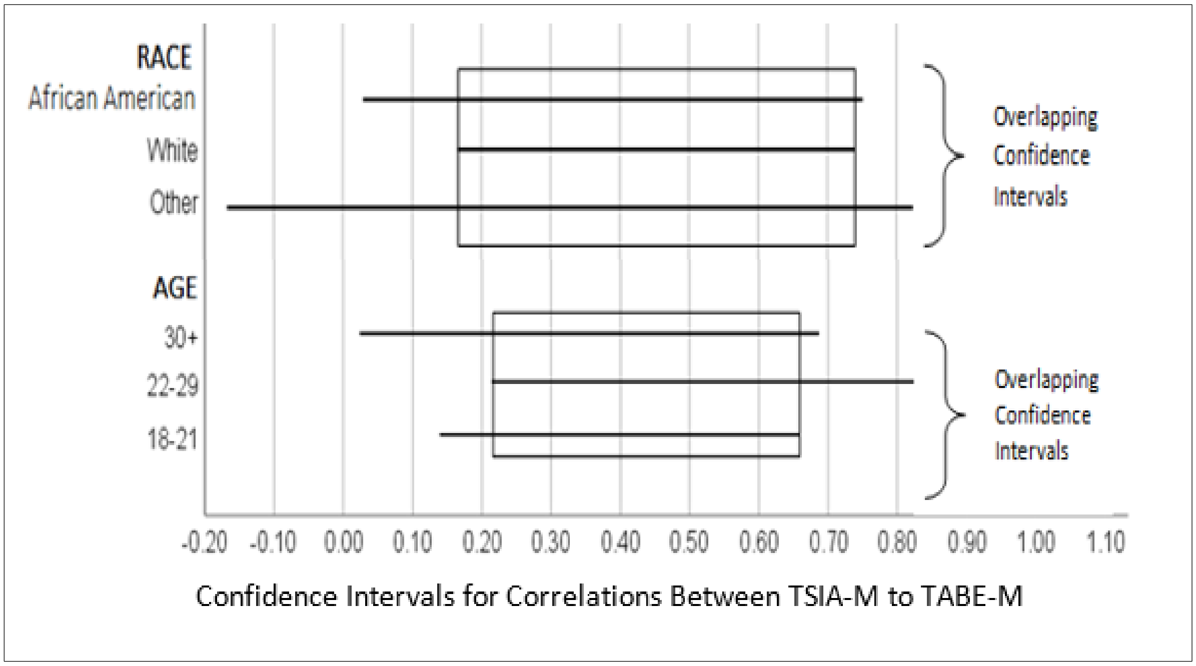

Figure 9. Confidence Intervals for TSIA-M to TABE-M Score Correlations

and Figure 9 illustrate overlapping confidence intervals across all demographic subgroups.

The Fisher's r-to-z transformation test results for Research Question 3 indicated that there were no statistically significant differences between test score correlations for females versus males or Hispanic versus non-Hispanic students. Calculating and plotting standardized confidence intervals yielded overlapping intervals between subgroups by race and by age group. The existence of overlapping confidence intervals signified that the correlations between subgroups based on race and subgroups based on age were not statistically significant. Therefore, I failed to reject the null hypothesis that there were no differences in the relationship between TSIA-M and TABE-M scores after controlling for student age, gender, race, and ethnicity.

Research Question and Null Hypothesis 4

- Does controlling for age, gender, race, or ethnicity change the relationships between NRS student placements on closely administered TSIA-M ABE Diagnostic and TABE-M assessments?
- H0: r_s= 0; there are no statistically significant differences in the relationships between NRS student placements on closely administered TSIA-M ABE Diagnostic and TABE-M assessments after controlling for age, gender, race, or ethnicity.

For this question, it was again necessary to determine differences between the relationships of demographic subgroup characteristics. However, instead of looking for changes in the relationship between scores on the assessments, I examined significant changes in the relationship between students' TSIA-M ABE Diagnostic and TABE-M NRS placement levels. As with Research Question 3, I used Fisher's *z* to compare the correlations of the subgroups for the dichotomous variables Gender and Ethnicity (Fisher, 1934). The results of the Fisher's *r-to z* transformation test, displayed in Table 18, indicated no significant correlation between the NRS levels from the mathematics TSIA ABE Diagnostic and the TABE-M between males and females or between Non-Hispanic and Hispanic students.

I used the same procedure outlined for Research Question 3 to determine the confidence intervals for the correlations for demographic subgroups within the race and age categories. Table 19 shows the lower and upper values of the confidence intervals for each subgroup correlation, and Figure 10 displays graphical representations of the confidence intervals for the test placement level correlations for Race and Age

Table 18

Significance of Correlations of TSIA-M ABE Diagnostic and TABE-M NRS Levels by Demographic Factor Using Fisher's r-to-z Transformation Test

Demographic Groups Compared	Fisher's z [a]
Female to Male	0.82
Non-Hispanic to Hispanic	1.61

Note. Z-values were calculated using an online calculator (Lowry, 2021). [a] *Z*-values greater than the absolute value of 1.96 indicate a statistically significant difference between the correlations of the two subgroups (Esri, n.d.).

variables. Both Table 19 and Figure 10 display overlapping confidence intervals across all Race and Age subgroups. Therefore, based on the Fisher's r-to-z transformation test for Gender and Ethnicity subgroups and overlapping confidence intervals for Race and Age subgroups, I failed to reject the null hypothesis that there are no statistically significant differences in the relationships between closely administered TSIA-M ABE Diagnostic and TABE-M NRS levels after controlling for age, gender, race, and ethnicity.

Chapter Summary

In this chapter, I first detailed the procedures I used to prepare the data set for analyses. Next, I provided descriptive statistics for the study variables TSIA-M scores, TSIA-M diagnostic ABE levels, TABE-M scores, TABE-M NRS levels, and the student demographic factors of age, gender, race, and ethnicity. I then described the statistical

Table 19

TSIA-M ABE Diagnostic and TABE-M NRS Level Correlations by Demographic Characteristics

Demographic Comparison	n	Spearman's r_s TSIA-M ABE Diagnostic Levels to TABE-M Levels	p	Fisher's z and Confidence Intervals
Age				
18–21	98	.334	.001	.35 [.14, .50]
22–29	25	.583	.002	.67 [.21, .81]
30 and over	29	.400	.032	.42 [.02, .68]
Race African American	41	.343	.030	.36 [.03, .59]
White	79	.378	.001	.40 [.67, .56]
Other	12	.793	.002	1.08 [.32, .95]

Note. Values in square brackets indicate the 95% confidence interval for each correlation.

analyses I used, their results, and whether or not I rejected or failed to reject each null hypothesis. Below I list the results of the null hypotheses testing:

1. H0: r_s= 0; there is no statistically significant relationship between student scores on closely administered mathematics TSIA-M and TABE-M assessments. I failed

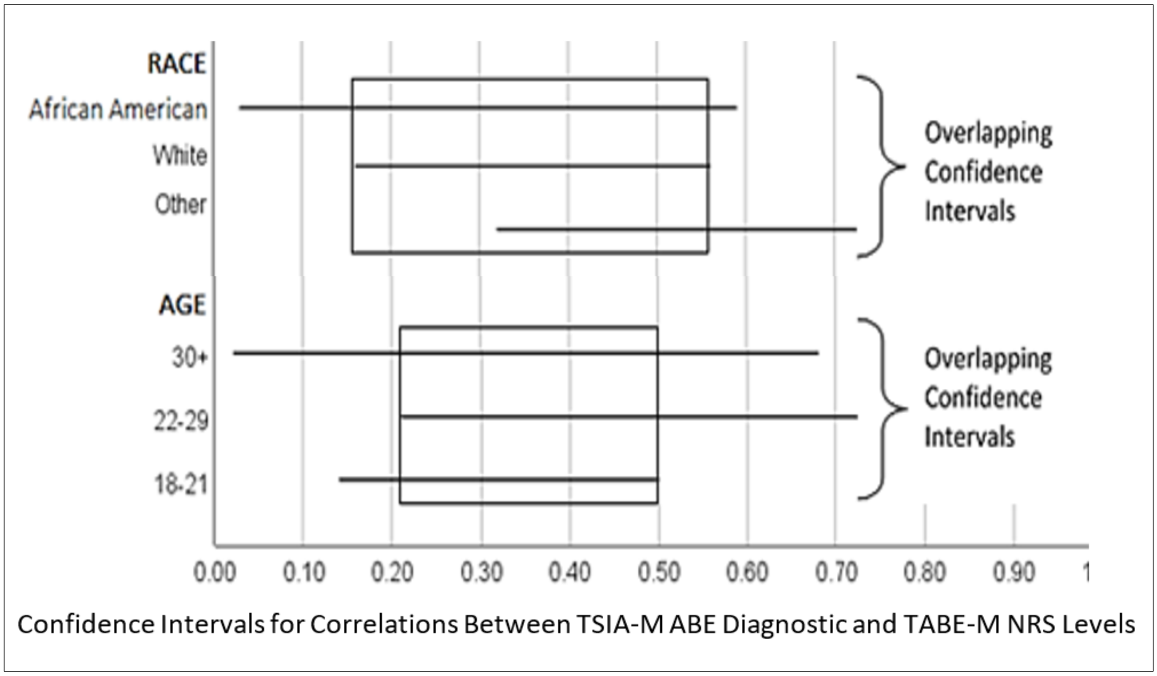

Figure 10. Confidence Intervals TSIA-M ABE Diagnostic to TABE-M Placement Correlations.

to reject Null Hypothesis 1 because I discovered a strong significant correlation between student scores on the TSIA-M ABE Diagnostic and the TABE-M.

2. H0: $r_s = 0$; there is no statistically significant relationship between NRS student placements from closely administered TSIA-M ABE Diagnostic and TABE-M assessments. I also rejected Null Hypothesis 2l because I found a significant moderate correlation between the NRS placement levels from the TSIA-M ABE Diagnostic and the TABE-M.
3. H0: $r_s = 0$; there are no statistically significant differences in the relationships between student scores on closely administered TSIA-M and TABE-M assessments after controlling for age, gender, race, or ethnicity. Because I discovered no significant differences between correlations of TSIA-M and TABE-

M scores across age, gender, race, or ethnicity student subgroups, I failed to reject Null Hypothesis 3.

4. H0: r_s= 0; there are no statistically significant differences in the relationships between NRS student placements on closely administered TSIA-M ABE Diagnostic and TABE-M assessments after controlling for age, gender, race, or ethnicity. I also failed to reject Null Hypothesis because I found no significant differences across student demographic subgroups between correlations of student NRS placements on the TABE-M ABE Diagnostic and the TABE-M.

In summary, the results indicated that (a) there are positive, moderate to strong statistically significant relationships between students' scores and placement levels on the two mathematics assessments; and (b) these relationships showed no statistically significant changes across demographic characteristics. In the following Chapter, I will discuss these results and their implications within the context of college placement in Texas.

V. DISCUSSION

The development and implementation of the Texas Success Initiative Assessment (TSIA) was an important component in Texas's efforts to ensure that students who were the least prepared academically for college would be placed into instructional services that would help them succeed. Texas required the test developer, the College Board, to incorporate a diagnostic module for students who scored below the minimum college readiness cut score. The diagnostic module intentionally included content at primary and secondary school academic levels ranging from first- to twelfth-grade. In addition to providing diagnostic information to help with advising and instruction, the module assigned the student an adult basic education (ABE) level from 1 to 6 that aligned to the educational functioning levels defined by the federal National Reporting System for Adult Education (NRS). Advisors placed students who scored a level 5 or 6 on the diagnostic into developmental education services at the college. Colleges had discretion whether they would serve students who were placed into ABE levels 1 through 4 the college's DE or continuing education programs or refer them to adult education services (THECB, 2014).

As a current practitioner working in a federally funded adult basic education (FF-ABE) program, I received students occasionally who had been referred due to TSIA placement levels of 1 through 4. Similar to the enrollment process for a community college, students in FF-ABE services must take a placement test. This assessment, known as the Tests of Adult Basic Education 11/12 (TABE 11/12), is an established standardized assessment used nationally and has undergone extensive piloting been normed through extensive piloting and the test publisher, Data Recognition Corporation (DRC). The

TABE 11/12's technical report is available on request from DRC. For its most recent version, the TABE 11/12, the test developers aligned the content and rigor to the federal adult education content standards, which in turn were developed to align with the Common Core State Standards for college readiness standards (Common Core States Initiative, 2018; Pimentel, 2013). Furthermore, the TABE is one of only two tests approved by the Department of Education for use in FF-ABE programs, and as such, must undergo an approval process in which it is evaluated using American Educational Research Association (AERA) et al.'s (1999) *Standards for Educational and Psychological Testing* (Adult Education and Family Literacy Act, 2011). Therefore, for the purposes of this investigation, I made the assumption that the TABE 11/12 placed students accurately into the six levels of the federal NRS measurement scale.

When I received the results of the TABE 11/12 scores and placement levels from students who had been referred to the FF-ABE program I worked with, I often noticed that the placement levels from the student's TSIA ABE diagnostic were often higher than their TABE 11/12's placement levels. Because both tests aligned their placement levels to the NRS's measurement scale for adult basic education proficiencies, I thought there might have been an issue with the TSIA's ABE diagnostic alignment. Informally, my colleagues and I began to assume that students would generally place lower on a TABE 11/12 than on the TSIA ABE diagnostic. When working with FF-ABE students who were interested in transitioning to postsecondary education, we believed that students who placed into a TABE 11/12 NRS level 4 were most likely going to score at DE levels 5 or 6 on the TSIA. In fact, we considered students who scored at NRS level 5 on the TABE as close to being college-ready and administered them GED practice tests immediately to screen for the student's likelihood of passing the GED high school equivalency test. My observations in the field led to my desire to undertake a formal investigation to explore the relationships between student performance and placement from on the TSIA and its ABE diagnostic level and the TABE 11/12.

Review of the Study

For this study, I used a non-experimental, correlational design to compare student scores on the TSIA mathematics (TSIA-M) and the TABE mathematics (TABE-M). I also compared TSIA mathematic ABE diagnostic (TSIA-M ABE diagnostic) placement levels to placement levels from the TABE-M. My worldview as a researcher was pragmatic. I also examined student relationships between the two tests by demographic characteristics to determine if there were any significant differences between demographic groups. As such, through my inquiry, I sought to understand the relationships between performance on the two assessments to see if a pattern emerged that would be useful to postsecondary and adult education practitioners.

My initial purpose was to find a practical solution to a real-world problem. However, as I began my investigation, I realized that the answers to my research questions would provide validation evidence in favor or against the use of the TSIA-M and TSIA-M ABE diagnostic for its use in placing students who are below college readiness levels. Because the answers to my research questions provide sources of evidence that can be used in a test validation process, I proposed a conceptual framework for test validation based on the AERA et al.'s (2014) *Standards for Educational and Psychological Testing* (see Chapter 1). It is through this framework that I conducted my literature review, chose my methodology, and developed my discussion.

Findings

In Dewey's (1938) theory of inquiry, the problematic phase does not end until the researcher is satisfied that enough evidence has been collected and tests completed to make warranted assertions about the object of the inquiry. As a pragmatist, I choose not to make warranted assertions because I know that truth is formed within the environment and society in which we interact. The purpose of the findings of an inquiry is meaningless unless they are useful either as guidance for future actions or as knowledge gained that can be built on through even more inquiry. My research questions were purposefully not set to prove or disprove specific hypothesis; rather, I devised them to allow flexibility in searching for warranted assertions about the TSIA-M ABE Diagnostic's role in assessment and placement of students who enter college with a need for basic skills instruction below the high school level. I present my research findings below within their relationship to my four research questions.

Research Question 1

The first two findings from my study come from my answer to Research Question 1: What is the relationship between student scores on closely administered TSIA-M and TABE-M assessments? The purpose of this question was to determine if the two tests were testing similar constructs, which is the case of these two assessments, was math skills at levels under college readiness proficiency. I also wanted to examine how well the TSIA-M placed students into the ABE Diagnostic component based on the DE cut score, which was set at 336.

Finding 1

My first finding was that the TSIA-M and the TSIA-M measured a very similar construct, which was math skills at levels below college readiness. First, I discovered that there was a strong positive correlation between student scores on the two assessments, meaning that students scored in similarly ranked positions on both assessments. According to my proposed *Standards* test validation framework in Chapter 1, one type of evidence for support of validation is an assessment's relation to an external variable. Specifically, this study used correlational analysis to locate evidence of convergent validity, which is the "extent to which two measures capture a common construct" (Carlson & Herdman, 2010, p. 18). The strong relationship I discovered between student scores on the TSIA-M and the TABE-M indicated that for students who scored between a 310 and a 335 on the TSIA-M, the TSIA-M measured the same underlying math proficiency skills as the TABE-M. Convergent validity builds support for an assessment's construct validity, and the strong correlation between the test scores supported the underlying construct of the TSIA-M of math proficiency under college readiness levels.

Finding 2

My second finding was that the TSIA-M's DE cut score of 336 performed well at predicting which students should have been administered the TSIA-M ABE Diagnostic. Ad hoc analysis of student TSIA-M ABE diagnostic placement levels revealed that none of the 152 students who were administered the TSIA-M ABE diagnostic over a 2-year period were placed into NRS level 6, and 12 students, or about 17%, were placed into NRS level 6. The purpose of the TSIA-M and its associated DE cut score of 336 is to place students who score below DE level (NRS levels 5 and 6) into the TSIA-M ABE

Diagnostic. Therefore, the accuracy rate for the TSIA-M's student placement into the TSIA-M ABE Diagnostic was quite high, at 83%. The strong convergent validity between the TSIA-M and the TABE-M and the high predictive ability of the TSIA-M for student placement into the TSIA-M ABE Diagnostic build evidence for validation of the use of the TSIA-M for determining which students needed to be administered the TSIA-M ABE Diagnostic.

Research Question 2

Two additional findings from my study come from the answer to Research Question 2: What is the relationship between student NRS levels on closely administered TSIA-M ABE Diagnostic and TABE-M assessments? My purpose for this question was to determine if the TSIA-M ABE Diagnostic and the TABE-M measured the same construct, which again was math proficiency at levels below college readiness. I also wanted to study the accuracy of the TSIA-M ABE Diagnostic's student placements into NRS levels.

Finding 3

The third finding from this study was that the TSIA-M ABE Diagnostic and the TABE-M measured somewhat similar constructs, which again was math proficiency under college-readiness levels. The results from the correlational analysis between the student placements levels on the two assessments revealed a significant correlation of moderate size, which again indicated the existence of convergent validity for the TSIA-M ABE Diagnostic. Convergent validity builds support for an assessment's construct validity, and the moderate correlation between the test placement levels supported the underlying construct of the TSIA-M of math proficiency under college readiness levels to some extent (Carlson & Herdman, 2010).

Finding 4

My fourth finding from my inquiry was that the TSIA-M ABE Diagnostic is not supported for its intended use of student placement into instruction under college-ready levels for students who placed into NRS levels 3 or 5. Ad hoc analysis determined that when the TABE-M score was used as a criterion, the TSIA-M ABE diagnostic correctly placed less than 15% of students, overplaced over 50% of students by one level, and overplaced almost 30% by two levels. As I discussed in the TABE 11/12 Instrumentation section in Chapter 3, the minimum consistency of classification between two tests that place students into the same levels should be at least 50%, with a higher level of consistency preferable (DRC, 2017). The consistency of classification analysis for NRS placement levels made by the TABE 11 and the TABE 12 averaged 66.2% across the five levels of math tests. The 15% agreement of the TSIA-M ABE diagnostic placement levels with the TABE-M placements levels is an extremely low level of placement accuracy.

However, when I analyzed the consistency of classification data at each NRS level, I discovered that between 50% and 55% of students who placed into NRS level would have still been placed correctly per THECB's guidelines. The THECB's recommendation for placing students with NRS levels 3 or 4 is co-enrollment into a DE class with co-enrollment in a basic adult skills education non-course-based option. Approximately 70% of students who took the TSIA-M ABE diagnostic were placed into NRS level 4, and approximately one half of these students were overplaced by one level.

This would have yielded a correct placement for approximately 35% of NRS level 4 students. Furthermore, the exact match in the consistency of classification for students with NRS level 4 placements was found to be 30% (see Table 14). This effectively raised the rate of correct course placement per THECB's guidelines to 55%, which is over the minimum 50% accuracy for consistency of classification recommended by DRC (2017).

Over the next 2 years, student TSIA scores will still be valid for placement. In order to generalize this finding on the acceptable placement accuracy for students who were placed into NRS level 4 by the TSIA-M ABE Diagnostic, I was able to compare the study's NRS placement level rates to statewide data. Daugherty et al. (2019b) provided percentages of students placed at each NRS level from 1 to 4 for fall 2015, as did Morgan (2019) for students who placed at these lower literacy levels for students in fall 2018. The high percentage of students placed into NRS level 4 remained similar between 2015 and 2020. In fall 2015, 79% of students statewide who placed into levels 1 to 4 were at NRS level 4, and in fall 2018, the figure was similar at 82% (Morgan, 2019). Using data from Table 14, I calculated that 136 students tested into NRS levels 1 to 4, with 78% of those students placing at NRS level 4.

Students who were placed into level 3 on the TSIA-M ABE Diagnostic, however, would not have been considered for referral to adult education services per THECB's guidelines. THECB recommends that if students have NRS levels of 1 or 2 in all three subject areas—math, reading, and writing—that they be considered for referral to adult education services. The accuracy of placement for these students was only 8.4%, and over 90% of students were overplaced. This included 56% who were overplaced by one level and 35% who were overplaced by two levels. Because my original data source from the study college provided me with TABE 11/12 reading, writing, and math scores, I was able to determine if any students had been placed by the TABE 11/12 into NRS levels of 1 or 2 in all subject areas. I discovered that 23 students, or 15%, would have met the criterion for referral to adult education services if they had been placed accurately by the TSIA-M ABE Diagnostic. Interestingly, none of these students had TSIA-M ABE Diagnostic placements of NRS level 2. Instead, these students had been overplaced one or two levels by TSIA ABE diagnostics.

Research Questions 3 and 4

My final finding comes from my answers to Research Questions 3 and 4: Does controlling for age, gender, race, or ethnicity change the relationships between student scores on closely administered mathematics TSIA-M and TABE-M assessments, and does controlling for age, gender, race, or ethnicity change the relationships between student NRS placements on closely administered mathematics TSIA-M ABE Diagnostic and TABE-M assessments? The purpose of these questions was to determine if the relationships between student performance scores and placements on the two assessments varied based on demographic characteristics. Therefore, I believed it was imperative to determine if the TSIA-M and TSIA-M ABE diagnostic advantaged or disadvantaged students based on their demographic characteristics.

Finding 5

My fifth finding was that there were no significant differences in the relationships between test scores or between test placement levels based on student gender, age group, race, or ethnicity. This finding suggests that the TSIA-M ABE effectively measures the mathematic skills of all test-takers, regardless of their demographic characteristic. A finding of no variance across subpopulations helps establish a source of evidence for consequential test use.

Implications

With the new state of Texas policy in effect as of fall 2021 mandating placement into corequisite DE for students with TSIA or TSIA 2.0 placement levels of 5 or 6, it is essential that students have adequate math skills in order to master the content of the DE and college-level math courses. Because the TSIA was retired on January 10, 2021 (19 Tex. Admin. Code §4.56), incoming students for fall 2021 were administered the TSIA 2.0 rather than the TSIA. However, if advisors relied on older TSIA ABE levels for placement purposes for fall 2021, students with TSIA-M ABE Diagnostic levels of 5 or 6 were most likely severely overplaced and would have been expected to master college-level math when their true math proficiency skills were estimated by the TABE-M as between fourth-and eighth-grade levels.

The second area of concern is the misplacement of students whose TSIA-M ABE Diagnostic placed them into level 3. The finding of the high likelihood that overplacement by one level would indicate that their true math proficiency is at second-to third-grade math skills. Whether or not adult education services are a good fit for students depends on several factors, but advisors who are simply following a placement chart would probably not discuss adult education services as an option. Furthermore, DE surveys of institutions indicated that many colleges were still referring students who placement levels of 1 to 4 into DE classes without co-enrollment into BASE support (THECB, 2018c).

I could locate no public THECB reports that indicated how many students in fall 2021 were placed using older TSIA ABE levels or how often advisors used older TSIA ABE levels in the future for placement purposes. According to state policy, TSIA scores and TSIA ABE diagnostic levels will remain valid for placement purposes for five years after the initial test date. It is worrisome that adults who delay college entry for 1 to 4 years after high school graduation may choose to use scores from the TSIA they were administered in high school rather than re-testing. According to a THECB presentation at the CASP 2021 statewide conference, THECB is aware that the TSIA overplaced the majority of students by one level. What is not clear is if THECB intends to issue guidance to colleges on how to use the inaccurate TSIA ABE diagnostic placement levels, especially when students were placed into level 3 or 5. The jump up to college-level coursework with corequisite DE rather than DE coursework with basic adult skills education (BASE) support could be from sixth grade to above college-level for close to half of the students with TSIA-determined placement levels and from fourth grade to above college-level for more than one-fourth of TSIA-placed level 5 students.

According to the *Standards*, the test developer is responsible for ensuring that a test is reliable and establishing its initial use based on standards-setting using content experts. However, with tests that rely on cut scores for placement into proficiency levels, concurrent and predictive validity is necessary to review the cut scores using empirical evidence from students' assessment scores and course outcomes. Unfortunately, it appears that the TSIA ABE Diagnostics were not included as part of the predictability study conducted by the College Board on behalf of THECB in 2017 (Cui & Bay, 2017). Although it is understandable that the focus for test validation be on cut scores and placements that affect the majority of college students, one of the reasons THECB developed the TSIA system to include diagnostics and placements into adult basic education skill levels was to assist students who entered college with skills below the ninth- to twelfth-grade skills taught in DE courses.

The College Board and THECB do not work directly with FF-ABE programs who would administer the TABE 11/12, so perhaps THECB was unaware that for two years--from January 2018 through January 2020—overplacement was occurring. However, the THECB recently stated that the original TSIA was overplacing students by about one level (Morales-Vale & Morgan, 2021). What is not clear is when this information became known or if the THECB intends to allow colleges to exempt students with TSIA placements of NRS level 5 from required placement into corequisite DE based on their realization that the TSIA ABE diagnostics placement levels inflate students' math proficiency. Furthermore, the College Board should review the data that is available from the Texas Workforce Commission to determine if students with TSIA-M ABE Diagnostic placements of NRS level 3 in all subject areas should be advised regarding adult education options.

Limitations

This investigation was limited in scope to the time frame, local geographic and educational setting, test administration processes, student effort during the test administration, student proficiency in math, and data collection and preparation. The TABE 11/12 became available for use in FF-ABE programs in Texas as of January 1, 2019, and the first version of the TSIA was phased out as of January 10, 2021. Therefore, closely timed administrations of both tests occurred only between January 1, 2019, and

January 10, 2021. A limitation to the study is the geographical and educational context of the research site, which was a medium-sized community college in southeast Texas. However, the population of the sample and of the college mirror the overall demographics of community college students in Texas.

The process used for testing and referral of students with TSIA-M ABE diagnostic scores was unique to the research site. The data set was one of opportunity; although there may be many instances of students who are administered the TSIA and the TABE 11/12 within a short time frame, the data systems for postsecondary IHEs and FF-ABE systems are separate, making it difficult to identify and obtain data. The findings are limited to students at the research site who took the TSIA and the TSIA-ABE diagnostic and agreed to take the TABE 11/12 and enroll in an upcoming College Success Academy. The efforts put forth by the student during the test administration are unknown but influence the validity of their scores. For example, students who did not complete the TABE 11/12 within a short time frame may simply be a different type of learner from students who did complete the TABE 11/12 in a timely manner. Furthermore, there may be differences between the type of student who was interesting in attending the CSA and therefore took the TABE 11/12.

Data collection was dependent on staff who worked at the college's testing center, institutional research center, and adult education department. Many students took the TSI multiple times. For purposes of this investigation, I only used the first attempt. However, due to the formatting of the data, extensive data preparation was necessary, which could have resulted in undetected errors.

Recommendations for Current Practice

My main desire in conducting this research was to determine if there was a way to use the TSIA-M ABE diagnostic information more valuable to FF-ABE staff and to college advisors who were using student's the TSIA-M ABE placements levels to make better-informed decisions for student enrollment in DE or other appropriate college services, including referral to a college-administered or externally administered FF-ABE program. FF-ABE practitioners working with students referred by the college to adult education services or working to transition adult education students to postsecondary were already becoming aware that the TSIA ABE NRS level placements were higher than the TABE 11/12 NRS placements simply through casual observation. However, this would most likely not happen in community college settings in which advisors do not work closely with FF-ABE program staff or have knowledge of TABE 11/12 NRS level placements.

College placement test scores are imperfect measures to predict future performance; therefore, many researchers have demonstrated that combining a student's placement test score with their high school grade point average (HSGPA) or high school transcript data improved placement accuracy (Barnett et al., 2020; Belfield & Crosta, 2012; Ngo & Kwon, 2015; Scott-Clayton, 2012; Scott-Clayton et al., 2014). However, not all adult learners have HSGPA or transcripts available, typically because they either earned a high school equivalency, completed high school in other countries, or have had many years pass since their high school graduation. Therefore, I separate my recommendations into two sections: one for students who have HSGPA available and the other for students who do not. Also, because any changes in placement may have unintended as well as intended consequences, it is important to vet these with other stakeholders involved in the placement process.

Students with High School Grade Point Average Available

Because there are only recommendations rather than mandates for placement or referral for students who present TSIA-M ABE diagnostic levels of 1 through 4, I make separate recommendations for these students than for students with levels 5 or 6. I recommend the following for college advisors for students with TSIA-M ABE diagnostic NRS levels of 1– 4:

- Consciously be aware of any implicit bias you may hold about student performance in math based on student race, gender, or ethnicity and strive to avoid applying stereotypes about a student's ability to achieve success based on any demographic characteristics (Coleman & Smith, 2021; Dadgar et al., 2021; Kosiewicz & Ngo, 2020; Park et al., 2018).
- Collect the following information from the student: HSGPA, date of high school graduation, full-time or part-time attendance status, any math-related military training or other math-related occupation training, and career and degree aspiration.
- Avoid placing students into corequisite DE unless the student has very strong evidence that supports high academic achievement, such as being is a recent high-school graduate with a strong HSGPA (Daugherty et al., 2019b).
- Research has shown that students who are less than two years out of high school and who attend full-time are more likely to perform better in their courses (Daugherty et al., 2019b; Stout, 2014). Inform students although they may be an

exception to these results research, but students who self-place should be made aware of this research. If necessary, explain that the TSIA ABE placement levels have been found by THECB to overplace students by about one level.

- Discuss all placement or referral options with the student and let the student decide on their placement (Coleman & Smith, 2021; Kosiewicz & Ngo, 2020; Park et al., 2018).
- If possible at the college or FF-ABE setting where the students are placed or referred, allow students to audit a second class or intervention for the first week or two so that they can decide if the audited course would be a better placement than the course in which they initially enrolled.
- Explain the process of class changes to the student and encourage them to initiate a class change within the time frame allowed by the college or FF-ABE program if they believe the placement is not in their best interest.

For students who have TSIA-M ABE diagnostic placement levels of 5 or 6, state policy requires them to be placed in corequisite DE. In lieu of any official THECB directive to adjust the use of TSIA-M ABE diagnostic NRS levels 5 or 6 for corequisite DE placement, I recommend the following for advisors:

- Consciously be aware of any implicit bias you may hold about student performance in math based on student race, gender, or ethnicity and strive to avoid applying stereotypes about a student’s ability to achieve success based on any demographic characteristics (Coleman & Smith, 2021; Dadgar et al., 2021; Kosiewicz & Ngo, 2020; Park et al., 2018).

- Explain to the student that they have the option of taking the TSIA 2.0. However, according to THECB, colleges must use the higher of the TSIA and TSIA 2.0 scores for placement purposes. Because the TSIA 2.0 has been reported to place students below what students were placed at with the original TSIA, students will most likely still be required to enroll in corequisite DE based on their TSIA-M ABE diagnostic placement levels (Morales-Vale & Morgan, 2021). Diagnostic information provided by the TSIA 2.0 might assist the advisor in recommendations as to the intensity of the corequisite DE course that supports the college gateway course.
- Collect the following information from the student: HSGPA, date of high school graduation, full-time or part-time attendance status, any math-related military training or other math-related occupation training, and career and degree aspiration.
- If math pathway choices are available at the college, place the student in a corequisite DE model with a gateway math course that aligns with their career and degree aspiration.
- If differentiated intensities of the corequisite DE class are available at the college (e.g., classes that meet either 1, 2, or 3 hours weekly), choose higher intensity DE courses for students who have been out of high school more than 2 years, have a HSGPA of C or below, and who will attend part-time in the highest intensity available. Students with a HSGPA of B or above who have been out of high school less than 2 years and who will attend full-time should be placed in DE courses with the least intensity. If the student wants additional justification for

their placement, explain that the TSIA ABE placement levels have been found by THECB to overplace students by about one level. As appropriate, inform students that research has shown that students who are less than two years out of high school and who attend full-time are more likely to perform better in their courses (Daugherty et al., 2019b; Stout, 2014).

- If possible and available at the college, allow students to audit a DE corequisite of a different intensity level for the first week or two so that they can compare two DE corequisite courses and make a class change to the audited course if desired within the time frame permitted by the college.
- Explain the process of class changes to the student and encourage them to initiate a class change within the time frame allowed by the college if they believe the intensity of the DE corequisite course is not the best placement for them.
- Ensure that students are aware of the academic support available to assist them in their gateway and DE math courses.

Students with No High School Grade Point Average Available

A subpopulation of adult learners will not have HSGPA available. Furthermore, effective placement for students using just the TSIA-M diagnostic NRS level may not be possible, even with the best intentions. For students with TSIA-M ABE diagnostic placement levels of 1–4, I recommend the same process as the one that I described for students with a HSGPA available but with the following changes:

- Ask open-ended questions about students' experiences in high school or other educational settings to determine if the student believed they were a high achieving student or if they struggled.

- Do not place students into corequisite DE unless the student has strong evidence of academic ability, such as a college degree in a math-related field from their own country or military or occupational training that is math-related (Daugherty et al., 2019b). The student should also be able to comfortably hold a simple conversation in English if they are an English language learner.

Placing students with TSIA-M ABE diagnostic placements levels of 5 or 6 is also similar to the process I described for students with HSGPA available. The only differences are the following:

- Ask open-ended questions about students' experiences in high school or other educational settings to determine if the student believed they were a high achieving student or if they struggled.
- If differentiated intensities of the corequisite DE class are available at the college (e.g., classes that meet either 1, 2, or 3 hours weekly), choose higher intensity DE courses for students who have been out of high school more than 2 years, have indicated that they struggled in school, and who will attend part-time. Students who will attend full-time and have a self-reported HSGPA of B or above, made a B or above in high school math courses, and took at least three years of math in high school should be placed in corequisite DE with the least intensity. Likewise, students with postsecondary degrees in math-related fields from another country should also be placed in corequisite DE with the least intensity if they can comfortably hold a simple conversation in English if they are an English language learner.

Recommendations for Future Research

My investigation of the relationship between student performance on the TSIA-M and the TABE-M and between student placement on the TSIA-M ABE diagnostic and the TABE-M is a small component within a much larger research agenda on the validation of college assessment and placement processes. Referencing my proposed *Standards* test validation framework from Chapter 1, sources of evidence for validation of a college placement and assessment system can focus on test content, on test-taker cognitive responses, on the test's internal consistency, on the test's relationships with other criteria, or on the consequences of test use.

Due to the high-stakes nature of college placement exams, it may be difficult for independent researchers to access data for research that examines test-taker cognitive responses and internal consistency. However, research undertaken to support the validation of test content could be more easily approached. For example, because students in high school often take several standardized exams, college placement test scores could be compared to student scores on one or more similar standardized assessments. Traditionally, predictive validity research based on test score relationships to an external variable or criterion has been common for the validation of assessment and placement procedures. Predictive validity studies that are longitudinal are needed in order to track students' short and long-term outcomes based on their initial college assessment and placement. Test publishers of the ACCUPLACER, COMPASS, and TSIA have determined the predictive validity of their math placement test scores on student gateway course grades as 65.8%, 68.8%, and 62.6%, respectively (see Table 1). However, additional research is needed to determine how changes to cut scores or the additional cognitive and non-cognitive measures, including student self-placement, can improve placement accuracy.

Tightly linked to predictive validity is research that empirically derives optimal cut scores for placement levels. This type of research must be done after a large number of test-takers have been administered the test, been placed, and have completed the course into which they were placed. As discussed in the Review of the Literature, performing statistical analyses on historical test scores and gateway course pass grades can determine optimal cut scores that would yield placement accuracy rates for gateway English that are as accurate as using a combination of test scores and HSGPA (Leeds & Mokher, 2020). Also, in many states, colleges within public IHE systems have the flexibility of setting cut scores as well as determining the criteria used for college placement (Hodges et al., 2020). Colleges or entire IHE systems can create customized placement systems based on data analytics techniques that can greatly increase placement accuracy (Bergman et al., 2021).

During my review of the literature, I noted that several researchers raised concerns about the gaps between college placement for students who are male, White, or Asian as compared to other demographic groups. Assessment and placement systems were also found to or predicted to advantage some student populations while disadvantaging others. I recommend continued research that extends the idea of test validation to placement validation. Studies that examine how changes in assessment and placement policies affect academic outcomes for different subpopulations of students are imperative to make informed decisions that promote equity for all students. Another theme that emerged in the literature was the role of advising in the student's placement decision; researchers have pointed to possible implicit bias and stereotyping based on student characteristics that might affect the advisor's recommendations. Therefore, I recommend that researchers focus on the role of advising in the placement process.

As part of the larger research agenda for increasing student success for all students through improvements in assessment and placements for incoming college students, the time is ripe to begin collecting data on the newly implemented TSIA 2.0. With the release of the new TSIA 2.0, there is a great opportunity to conduct independent research that collects evidence for validity. I challenge future researchers to focus their research on the students who the TSIA 2.0 placed in levels 1 through 4. To my knowledge, no information has been publicly released that discusses research on the concurrent or predictive validity of the TSIA 2.0 diagnostic's cut scores or placement levels. Although THECB has scheduled a predictive validity study from the College Board for 2022, I am unaware if the College Board will include the TSIA 2.0 diagnostic component as part of that study (Morales-Vale & Morgan, 2021).

Finally, there is a need for research that collects and analyzes the voices of the students. Although I located two studies that investigated faculty perceptions on the TSAI as a placement tool, I did not locate any research that looked at TSIA testing from the student's perspective. Although quantitative research is imperative to discover if what we as educators do is beneficial, it is the qualitative research that lets us understand why or why not what we do is effecting positive change.

Conclusions

As a pragmatic researcher, my approach to this investigation has been one that follows Dewey's (1938) theory of inquiry. I entered the inquiry based on a real-world puzzle that I wanted to solve regarding the misalignment I observed as a practitioner between two assessments that were both designed to measure proficiency in elementary through high school mathematics and were both designed to place students based on the National Reporting System of Adult Education's adult education proficiency level descriptors (DRC, 2017; THECB, 2014).

The intention of diagnostic assessment and ABE placement levels must be applauded. Few states have instituted programming to assist the small percentage of learners who test below high school academic skills levels. However, issues with student overplacement into NRS adult basic education skills level has affected the efficacy of THECB recommendation for student placement into basic adult skills education and developmental education instruction offered at Texas IHEs and referral for students to partnering adult education services.

By establishing significant correlations between student test scores and placements, I have added to the construct validity argument for the use of the TSIA-M to measure mathematics proficiency at levels below college readiness. However, I have also established that the TSIA-M ABE diagnostic consistently overplaced over 80% of all students by at least one level. Placement levels of 3 or 5, if interpreted as accurate proxies for student math proficiency, could lead to advisor placement according to THECB recommendation that would, in fact, severely overplace the student. Awareness of the high level of overplacement on the TSIA-M ABE Diagnostic and the nuances in the relationship between the assigned score and the probable overplacement can assist colleges with updating their placement processes. This information could also be useful to FF-ABE working to transitions students to postsecondary and college advisors who work to appropriate place students in basic skills instruction.

Ultimately, “scale scores, proficiency levels, and cut scores can be central to the use and interpretation of test scores. For that reason, their defensibility is an important consideration in test score validation for the intended purposes” (AERA et al., 2014, p. 95). Based on this study’s findings on the consistency of classification demonstrating overplacement by the TSIA-M ABE diagnostic and on THECB’s public announcement that compared to the TSIA 2.0 diagnostic, the TSIA ABE diagnostics overplaced most students by approximately one level (Morales-Vale & Morgan, 2021), placement levels assigned by the TSIA-M ABE Diagnostic, in particular for levels 3 and 5, may not defensible for their past and current use as college placement instruments.

REFERENCES

ACT. (2020). ACT Technical Manual. https://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf

Adult Education and Family Literacy Act, 34 C.F.R. § 462.13. (2011). https://www.govinfo.gov/content/pkg/CFR-2011-title34-vol3/xml/CFR-2011-title34-vol3-sec462-13.xml

Aguinis, H., Culpepper, S. A., & Pierce, C. S. (2016). Differential prediction generalization in college admissions testing. *Journal of Education Psychology, 108(*7), 1045–1059. https://doi.org/10.1037/edu0000104

American Educational Research Association. (2000). *Position statement on high-stakes testing.* https://www.aera.net/About-AERA/AERA-Rules-Policies/Association-Policies/Position-Statement-on-High-Stakes-Testing

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* https://www.aera.net/Portals/38/1999%20Standards_revised.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

Arendale, D. R. (2005). Terms of endearment: Words that define and guide developmental education. *Journal of College Reading and Learning, 35*(2), 66–82. https://doi.org/10.1080/10790195.2005.10850174

Arendale, D. R. (2011). Then and now: The early years of developmental education. *Research and Teaching in Developmental Education, 18*(2), 5–23. https://www.jstor.org/stable/42802530

Atkinson, R., & Geiser, S. (2009). *Reflections on a century of college admissions tests* (Research & Occasional Paper Series: CSHE.4.09). Center for Studies in Higher Education: University of California, Berkeley. https://cshe.berkeley.edu/sites/default/files/publications/rops-atkinsongeiser-tests-04-15-09.pdf

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

Bahr, P. R. (2010). Revisiting the efficacy of postsecondary remediation: The moderating effects of depth/breadth of deficiency. *The Review of Higher Education, 33*(2), 177–205. http://doi.org/10.1353/rhe.0.0128

Bahr, P. R., Fagioli, L. P., Hetts, J., Hayward, C., Willett, T., Lamoree, D., Newell, M., Sorey, K., & Baker, R. B. (2019). Improving placement accuracy in California's community colleges using multiple measures of high school achievement. *Community College Review, 47*(2), 178–211. https://doi.org.10.1177/0091552119840705

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

Barnett, E., Bergman, P., Kopko, E., Reddy, V., Belfield, C., & Roy, S. (2018). *Multiple measures placement using data analytics: An implementation and early impacts report.* Center for the Analysis of Postsecondary Readiness. https://postsecondaryreadiness.org/wp-content/uploads/2018/09/multiple-measures-placement-using-data-analytics.pdf

Barnett, E., Kopko, E., Cullinan, D., & Belfield, C. R. (2020). *Who should take college-level courses? Impact findings from an evaluation of a multiple measures assessment strategy.* Center for the Analysis of Postsecondary Readiness. https://postsecondaryreadiness.org/dev/wp-content/uploads/2020/10/multiple-measures-assessment-impact-findings.pdf

Barnett, E., & Reddy, V. T. (2017). *College placement strategies: Evolving considerations and practices* (Working Paper). Center for the Analysis of Postsecondary Readiness. https://www.luminafoundation.org/wp-content/uploads/2017/08/college-placement-strategies.pdf

Bergman, P., Kopko, E., Rodriguez, J. E. (2021). *Using predictive analytics to track students: Evidence from a seven-college experiment* (NBER Working Paper 28948). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28948/w28948.pdf

Belfield, C. R. (2014). *Improving assessment and placement at your college: A tool for institutional researchers.* Community College Research Center. https://ccrc.tc.columbia.edu/media/k2/attachments/improving-assessment-placement-institutional-research.pdf

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

Bettinger, E. P., Evans, B. J., & Pope, D. G. (2011). *Improving college performance and retention the easy way: Unpacking the ACT exam* (NBER Working Paper 17119). National Bureau of Economic Research. https://doi.org/10.3386/w17119

Bleasby, J. (2011). Overcoming Relativism and Absolutism: Dewey's ideals of truth and meaning in philosophy for children. *Educational Philosophy and Theory, 43*(5), 453–466. http://doi.org/10.1111/j.1469-5812.2009.00567.x

Boatman, A., & Long, B. T. (2018). Does remediation work for all students? How the effects of postsecondary remedial and developmental courses vary by level of academic preparation. *Educational Evaluation and Policy Analysis, 40*(1), 29–58. http://doi.org/10.3102/0162373717715708

Boylan, H. R., & Bonham, B. S. (2007). 30 years of developmental education: A retrospective. *Journal of Developmental Education, 30*(3), 2–4. https://www.jstor.org/stable/42775241

Boylan, H. R., & Bonham, B. S. (2011). Seven myths about developmental education. *Research & Teaching in Developmental Education, 27*(2), 29–36. https://www.jstor.org/stable/42802368

Boylan, H. R., Calderwood, B. Levine-Brown, P. & Anthony, S. W. (2017). 40th anniversary research timeline. *Journal of Developmental Education, 40*(3), 2–6. https://www-jstor-org.libproxy.txstate.edu/stable/44987453

Bracco, K. R., Dadgar, M., Austin, K., Klarin, B., Broek, M., Finkelstein, N., Mundry, S., & Bugler, D. (2014). *Exploring the use of multiple measures for placement into college-level courses: Seeking alternatives or improvements to the use of a single standardized test.* WestEd. https://www.wested.org/wp-content/uploads/2016/11/1397164696product55812B-3.pdf

Brown, M. J. (2012). John Dewey's logic of science. *The Journal of the International Society for the History of Philosophy of Science, 2(*2), 258–306. https://www.jstor.org/stable/10.1086/666843

Buckingham, B. R. (1921). Intelligence and its measurement: A symposium XIV. *Journal of Educational Psychology, 12*(5), 271–275. https://doi.org/10.1037/h0066019

Calcagno, J. C., & Long, B. T. (2009). *Evaluating the impact of remedial education in Florida.* (ED506374) ERIC. https://eric.ed.gov/?id=ED506374

Camara, W., Allen, J., & Moore, J. (2017). Empirically based college-and career-readiness cut scores and performance standards. In K. L. McClarty, K. D. Mattern, & M. N. Gaertner (Eds.), *Preparing students for college and careers: Theory, measurement, and educational practice* (pp. 70–81). Routledge.

Capps, J. (2019). The pragmatic theory of truth. In E. N. Zalta (Ed.) The Stanford Encyclopedia of Philosophy (Summer 2019 Edition). https://plato.stanford.edu/archives/sum2019/entries/truth-pragmatic

Carlson, K., & Herdman, A. (2010). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, *15*(1), 17–32. https://doi.org/10.1177/1094428110392383

Carter, A. F. (2019). *Key to success? Relationship between Texas success initiative assessments and course performance in first-year, credit-bearing English and math classes* [Doctoral dissertation, University of Texas at Arlington]. UTA ResearchCommons. https://rc.library.uta.edu/uta-ir/bitstream/handle/10106/28574/CARTER-DISSERTATION-2019.pdf?sequence=1&isAllowed=y

Castro, E. L. (2013). Racialized readiness for college and career: Toward an equity-grounded social science of intervention programming. *Community College Review, 41*(4), 292–310. http://doi.org/10.1177/0091552113504291

Chen, X., & Simone, S. (2016). *Remedial coursetaking at U.S. public 2- and 4-year institutions: Scope, experiences, and outcomes* (NCES 2016–405). U.S. Department of Education. https://nces.ed.gov/pubs2016/2016405.pdf

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). Lawrence Erlbaum Associates.

Coleman, D., & Smith, D. (2021). Beyond predictive validity: A mixed method study of self-directed developmental education placement at a small community college. *Community College Journal of Research and Practice, 45*(6), 403–422. http://doi.org/10.1080/10668926.2020.1719938

College Board. (2014). *TSIA assessments program manual combined*. https://secure-media.collegeboard.org/digitalServices/pdf/accuplacer/tsi-assessment-manual-december-2014.pdf

College Board. (2018a). *ACCUPLACER program manual.* https://accuplacer.collegeboard.org/accuplacer/pdf/accuplacer-program-manual.pdf

College Board. (2018b). *Texas Success Initiative. Interpreting your scores.* https://accuplacer.collegeboard.org/accuplacer/pdf/tsia-interpreting-your-score-brochure.pdf

College Board. (2021). *ACCUPLACER® placement validity study guide.* https://aces.collegeboard.org/pdf/aces-accuplacer-placement-validity-study-guide.pdf

Common Core State Standards Initiative. (n.d.). Read the Standards. http://www.corestandards.org/read-the-standards/

Common Core State Standards Initiative. (2018). *Development process*. http://www.corestandards.org/about-the-standards/development-process/#adoption

Complete College America. (n.d.). *Data Dashboard.* https://completecollege.org/data-dashboard

Complete College America. (2017). *Common college completion metrics technical guide.* https://completecollege.org/wpcontent/uploads/2017/08/2017MetricsTechnicalGuide.pdf

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education 49*, 560–575. https://doi.org/10.1111/medu.12678

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* SAGE.

Crisp, G., & Delgado, C. (2014). The impact of developmental education on community college persistence and vertical transfer. *Community College Review, 42* (2), 99–117. http://doi.org/10.1177/0091552113516488

Cronbach, L. J., & Meehle, P. E. (1955). Construct validity in psychological tests. *Psychology Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957

Cross, P. (1976). *Accent on learning. improving instruction and reshaping the curriculum*. Jossey-Bass.

CTB/McGraw-Hill. (2004). *TABE 9 and 10 technical report.* Author.

Cui, Z., & Bay, L. (2017). *The predictive placement validity study of Texas Success Initiative assessment.* http://www.thecb.state.tx.us/reports/PDF/9567.PDFCFID=62923641&CFTOKEN=86303321

Dadgar, M., Buck, D., & Burdman, P. (2021). *Solving for equity: Design and implementation of new postsecondary math pathways*. Just Equations. https://justequations.org/wp-content/uploads/Just-Equations-2021-Report-Solving-for-Equity-v4.pdf

Data Recognition Corporation. (2017). *TABE 11/12 technical report.* Author.

Data Recognition Corporation. (2019). *TABE Tests of Adult Basic Education.* https://tabetest.com

Daugherty, L., Gerber, R., & Martorell, F., Miller, T., & Weisburst, E. (2021). Heterogeneity in the effects of college course placement. *Research in Higher Education 62*, 1086–1111. https://doi.org/10.1007/s11162-021-09630-2

Daugherty, L., Miller, T., & Carew, D. G. (2019a). *Supporting college enrollees who test at the lowest levels of readiness: Lessons from Texas community colleges* (RR-2938–IES). RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RR2900/RR2938/RAND_RR2938.pdf

Daugherty, L., Miller, T., & Carew, D. G. (2019b). *Supporting college enrollees who test at the lowest levels of readiness: Lessons from Texas community colleges — Technical Appendix.* https://www.rand.org/content/dam/rand/pubs/research_reports/RR2900/RR2938/RAND_RR2938z1.appendix.pdf

Desai, S. A. (2012). Is comprehensiveness taking its toll on community colleges? An in-depth analysis of community colleges' missions and their effectiveness. *Community College Journal of Research and Practice, 36*(2), 111–121. http://doi.org/10.1080/10668920802611211

Dewey, J. (1910). *How we think*. D. C. Heath & Co. https://www.scribd.com/read/372269597/How-We-Think#

Dewey, J. (1931). *Philosophy and civilization*. G. P. Putnam's Sons.

Dewey, J. (1938). *Logic: The Theory of Inquiry*. Henry and Holt and Company.

Education Commission of the States. (2017). *Fifty state-comparison: Does the policy include multiple measures to determine college readiness?* http://ecs.force.com/mbdata/MBquest3RTA?Rep=SA1702

Ellis, S. F. (n.d.). *Correlation between various placement instruments for reading, language/writing, mathematics, elementary algebra.* https://wvde.state.wv.us/abe/documents/CorrelationBetweenVariousPlacementInstruments.pdf

Esri. (n.d.). *What is a z–score? What is a p–value?* https://pro.arcgis.com/en/pro–app/latest/tool–reference/spatial–statistics/what–is–a–z–score–what–is–a–p–value.htm

Fain, P. (2015, June 18). Finding a new compass. Inside Higher Education. https://www.insidehighered.com/news/2015/06/18/act-drops-popular-compass-placement-test-acknowledging-its-predictive-limits

Fields, R., & Parsad, B. (2012). *Tests and cut scores used for student placement in postsecondary education: Fall 2011*. National Assessment Governing Board. https://eric.ed.gov/?id=ED539918

Fike, D., & Fike, R. (2008). Predictors of first-year student retention in the community college. *Community College Review, 36,* 68–88. http://doi.org/10.1177/0091552108320222

Fisher, R. A. (1934*). Statistical Methods for Research Workers* (5th ed.) Oliver and Boyd, Ltd. http://www.haghish.com/resources/materials/Statistical_Methods_for_Research_Workers.pdf

Fong, K. E., Melguizo, T., Prather, G. (2015). Increasing success rates in developmental math: the complementary role of individual and institutional characteristics. *Research in Higher Education, 56*, 719–749. https://doi.org/10.1007/s11162-015-9368-9

Gardner, M. J., & Altman, D.G. (2000). Estimating with confidence. In D. G. Altman, D. Machin, T. N. Bryant, & M. J. Gardner (Eds.) *Statistics with Confidence* (2nd ed., pp. 3–5). British Medical Journal.

GED Testing Service. (2017). *March 2017 efficacy research report.* Author. https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/efficacy-and-research/reports/GED-Testing-Research-Report-v8.pdf

GED Testing Service. (2019a). Can you go to college with a GED? GED Blog for Students. https://ged.com/blog/can-you-go-to-college-with-a-ged

GED Testing Service. (2019b). *What is the GED?* https://ged.com/blog/what-is-a-ged

Gignac, G.E. (n.d.). *Spearman Correlation Confidence Intervals.sps.* https://tinyurl.com/y4e7dtvn

Gignac, G.E. (2019a). Bivariate correlation. In Author (Ed.), *How2StatsBook online* (pp. C3.1– C3-41). https://drive.google.com/file/d/1Uajjp8NrDGyYdIxekr5JWP2pT5kBPcE3/view

Gignac, G.E. (2019b). *V5.18 - Spearman Correlation Confidence Intervals SPSS* (Video). YouTube. https://www.youtube.com/watch?v=gKHgCq5E864

Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems, 21*(2), 135–146. http://doi.org:10.1057/ejis.2011.54

Goudas, A. M. (2019*). Multiple measures for college placement: Good theory, poor implementation* (Working Paper No. 6). Community College Data. http://communitycollegedata.com/articles/multiple-measures-for-college-placement

Green, A. (2019). *Postsecondary enrollment type 1b MSG* [Webinar]. https://twc.adobeconnect.com/_a952639191/p7lew8evk73v/?proto=true

Gutek, G. L. (2012). Introduction. In Author (Ed.), John Dewy: *How we think* (pp. 4–14). https://www.scribd.com/read/431784006/How-We-Think-Barnes-Noble-Digital-Library#

Higher Education Act of 1965, as Amended, 34 C.F.R. § 600.2. (2019). https://www.govinfo.gov/content/pkg/CFR-2019-title34-vol3/xml/CFR-2019-title34-vol3-part600.xml#seqnum600.2

Hodara, M., & Cox, M. (2016). *Developmental Education and College Readiness at the University of Alaska* (REL 2016–123). US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest. https://ies.ed.gov/ncee/edlabs/regions/northwest/pdf/REL_2016123.pdf

Hodara, M., Jaggars, S. S., & Karp, M. M. (2012). *Improving developmental education assessment and placement: Lessons from community colleges across the country* (CCRC Working Paper No. 51). Community College Research Center. http://eric.ed.gov/?id=ED537433

Hodges, R., McConnell, M., Lollar, J., Guckert, D. A., Owens, S., Gonzales, C., Wu, N., & Shinn, H, B. (2020). Developmental education policy and reforms: A 50-state snapshot. *Journal of Developmental Education, 44*(1), 2–17.

Holschuh, J. P., & Paulson, E. (2013). *The terrain of college developmental reading*. College Reading & Learning Association. http://www.crla.net/publications.htm

Horn, L., & Nevill, S. (2006). *Profile of undergraduates in U.S. postsecondary education institutions, 2003–04: With a special analysis of community college students* (NCES 2006–184). NCES. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006184

Hu, S., Park, T. J., Woods, C. S., Tandberg, D. A., Richard, K., & Hankerson, D. (2016). *Investigating developmental and college-level course enrollment and passing before and after Florida's developmental education reform* (REL 2017–203). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://ies.ed.gov/ncee/edlabs

Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research, 103,* 219–230. https://doi.org/10.1007/s11205-011-9843-4

Hughes, K. L., & Scott-Clayton, J. (2011). *Assessing developmental assessment in community colleges: A review of the literature* (CCRC Working Paper No. 19). https://ccrc.tc.columbia.edu/media/k2/attachments/assessing-developmental-assessment.pdf

Jacobs, J., & Tolbert-Bynum, P. (2008). *Shifting gears: Community colleges and adult basic education*. Community College Research Center. https://ccrc.tc.columbia.edu/media/k2/attachments/shifting-gears-adult-education.pdf

Jaggars, S. S., & Stacey, G. W. (2014). *What we know about developmental education outcomes* (CCRC Research Overview). Community College Research Center. http://academiccommons.columbia.edu/catalog/ac:170486

Johnson, M. (2019). *A new TABE for a new era* [PowerPoint slides]. https://www.aeaut.org/documents/conference/2019_presentations/Capra%201%20-%20Michael%20Johnson%20-%20NEW%20TABE%202019%20for%20Utah%209-27-19.pptx

Kane, M. T. (2000). *Current Concerns in Validity Theory* (ED446094). ERIC. https://eric-ed-gov.libproxy.txstate.edu/?id=ED446094

Kane, M. T. (2008). Terminology, emphasis, and utility in validation. Educational Researcher, 37(2), 76–82. https://doi.org/10.3102/0013189X08315390

Kane, M. T. (2013). Validation as a pragmatic, scientific activity. *Journal of Educational Measurement, 50*(1), 115–122. https://doi.org/10.1111/jedm.12007

Kelley, T. L. (1927). Interpretation of educational measurements. In L. M Terman (Ed.) *Measurement and Adjustment Series* (pp. 1–386). World Book. https://babel.hathitrust.org/cgi/pt?id=mdp.39015001994071&view=1up&seq=5

Kosiewicz, H., & Ngo, F. (2020). Giving community college students choice: The impact of self-placement in math courses. *American Educational Research Journal, 57*(3), 1358–1391. https://doi.org/10.3102/0002831219872500

Lee, B. (2018). *Texas Success Initiative Test Scores as a Predictor of College Mathematics Success* (Doctoral dissertation, Walden University).

Leeds, D. M., & Mokher, C. G. (2020). Improving indicators of college readiness: methods for optimally placing students into multiple levels of postsecondary coursework. *Educational Evaluation and Policy Analysis 42*(1), 87–109. http://doi.org/10.3102/0162373719885648

Lesik, S. (2007). Do developmental mathematics programs have a causal impact on student retention? An application of discrete-time survival and regression-discontinuity analysis. *Research in Higher Education, 48*(5), 566–591. http://doi.org/10.1007/s11162-006-9036-1

Lissitz, R., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher, 36*, 437–448. https://doi.org/10.3102/0013189X07311286

Literacy Information and Communication System. (n.d.). *College and Career Readiness Standards for Adult Education*. https://lincs.ed.gov/professional-development/resource-collections/profile-521

Lowry, R. (2021). *Significance of the difference between two correlation coefficients*. Vassarstats. http://vassarstats.net/rdiff.html

Martorell, P., & McFarlin, I., Jr. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *Review of Economics and Statistics, 93*(2), 436–454. http://doi.org/10.1162/REST_a_00098

Mattern, K. D., & Packman, S. (2009). *Predictive validity of ACCUPLACER™™ scores for course placement: A meta-analysis* (Research Report No. 2009-2). (ED561046). ERIC. https://eric.ed.gov/?id=ED561046

Mattern, K. D., & Patterson, B. F. (2013). Test of slope and intercept bias in college admissions: A response to Aguinis, Culpepper, and Pierce (2010*). Journal of Applied Psychology, 98*(1), 134–147. http://doi.org/10.1037/a0030610

Maxwell, L. E. (2021). *Predictors of student success in a gateway mathematics course: A Correlational Study* (Publication 28544983) [Doctoral dissertation, Grand Canyon University]. Proquest Dissertations and Theses Global. https://www.proquest.com/docview/2551571193/7E60686BBBF94922PQ/1?accountid=5683/

McAdams, F. C. (2017). *Developmental education faculty perspective of the Texas Success Initiative assessment as a placement exam* (Publication 10756384) [Doctoral dissertation]. https://www.proquest.com/docview/2013215101?accountid=5683

McClarty, K. L. Loomis, S. C., & Pitoniak, M. J. (2017). Building external validity into the process: Evidence-based readiness standards. In K. L. McClarty, K. D. Mattern, & M. N. Gaertner (Eds.). *Preparing students for college and careers: Theory, measurement, and educational practice.* https://doi.org/10.4324/9781315621975

McClarty, K. L., Way, W. D., Porter, A. C., Beimers, J. N., & Miles, J. A. (2013). Evidence-Based standard setting: Establishing a validity framework for cut scores. *Educational Researcher, 42*(2), 78–88. https://doi.org/10.3102/0013189X12470855

Medhanie, A. G., Dupuis, D. N., LeBeau, B., Harwell, M. R., & Post, T. R. (2012). The role of the ACCUPLACER mathematics placement test on a student's first college mathematics course. E*ducational and Psychological Measurement, 72*, 332–351. http://doi.org/0.1177/0013164411417620

Melguizo, T., Bos. L., Ngo, F., Mills, N., & Prather, G. (2016). Using a regression discontinuity design to estimate the impact of placement decisions in developmental mathematics. *Research in Higher Education 57*, 123–151. https://doi.org/10.1007/s11162-015-9382-y

Melguizo, T., Hagedorn, L. S., & Cypers, S. (2008). The need for remedial/developmental education and the cost of community college transfer: Calculations from a sample of California community college transfers. *Review of Higher Education, 31*, 401–431. http://doi.org/10.1353/jhe.2014.0025

Melguizo, T., Prather, G., & Bos, J. (2013). *Are community colleges making good placement decisions in their math trajectories?* University of Southern California Rossier School of Education. https://pullias.usc.edu/wp-content/uploads/2013/10/Good_Placement_Decisions_Brief.pdf

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). Macmillan Publishing, American Council on Education.

Messick, S. (1994). *Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning*. (Research Report RR-94-45) (ED380496). ERIC. https://eric-ed-gov.libproxy.txstate.edu/?id=ED380496

Mislevy, R. J. (2009). *Validity from the perspective of model-based reasoning* (CRESST Report 752). (ED507085). ERIC. https://files.eric.ed.gov/fulltext/ED507085.pdf

Monsaas, J. A. (2007). Test review of Tests of Adult Basic Education, Forms 9 & 10. In K. F. Geisinger, R. A. Spies, J. F. Carlson, & B. S. Plake (Eds.), *The seventeenth mental measurements yearbook* (17:187*)*. Buros Center for Testing.

Morales-Vale, S., & Daniels, T. (2015, March 6). *The 2015 THECB updates: TSIA, developmental and adult education.* http://www.tabphe.org/Resources/Documents/TSI%20Developmental%20and%20Adult%20Education.pdf

Morales-Vale, S. & Montognese, K. (2017). *College readiness and TSI updates. Early college high school conference* [PowerPoint slides]. https://www.texasccrsm org/ECHS/Benchmark-5-Academic-Rigor-and-Readiness/Resources/Exemplar-Campus-Practices/SMorales_Vale_TSI_CollegeReadiness

Morales-Vale, S., & Morgan, K. (2021, October). *THECB townhall meeting.* [PowerPoint slides]. College Academy Success Programs 2021 Conference. https://tinyurl.com/THECBTOWNHALL

Morgan, D. L., & Michaelides, M. P. (2005). *Setting cut scores for college placement* (Research Report No. 2005-9). (ED562865). ERIC. https://files.eric.ed.gov/fulltext/ED562865.pdf

Morgan, K. (2019). *The future of serving unprepared students in the state of Texas* [PowerPoint slides]. https://www.catchthenext.org/The-Future-of-Texas-DE-CTN-Webinar-4-18-19-1.pdf

Moss, B. G., Yeaton, W., & Lloyd, J. E. (2014). Evaluating the effectiveness of developmental mathematics by embedding a randomized experiment within a regression discontinuity design article. *Educational Evaluation and Policy Analysis, 36*(2), 170–185. http://doi.org/10.3102/0162373713504988

National Association for College Admission Counseling. (2016). *Use of predictive validity studies.* https://www.nacacnet.org/globalassets/documents/publications/research/testvalidity.pdf

National Center for Education Statistics. (n.d.). *Beginning postsecondary students longitudinal study* (BPS:12/17). U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/datalab/powerstats/71-beginning-postsecondary-students-2012-2017

National Center for Public Policy and Higher Education, & Southern Regional Board. (2010). *Beyond the rhetoric: Improving college readiness through coherent state policy*. https://www.sreb.org/sites/main/files/file-attachments/beyond_the_rhetoric.pdf?1475793547

National Reporting System for Adult Education. (2012a). *National Reporting System for Adult Education (NRS) 2012 changes: Automatic cohort designation and new measures frequently asked questions.* U.S. Department of Education. https://cdn.education.ne.gov/wp-content/uploads/2017/07/FAQs20about20the20NRS20changes.3.30.12202.pdf

National Reporting System for Adult Education. (2012b). *Table 4: Educational gains and attendance by educational functioning level.* U.S. Department of Education. https://nrs.ed.gov/rt/tx/2012/table-4

National Reporting System for Adult Education. (2013). *Table 4: Educational gains and attendance by educational functioning level*. U.S. Department of Education. https://nrs.ed.gov/rt/tx/2013/table-4

National Reporting System for Adult Education. (2015a). *Table 1: Participants by entering educational functioning level, ethnicity, and sex.* U.S. Department of Education. https://nrs.ed.gov/rt/tx/2015/table-1

National Reporting System for Adult Education. (2015b). *Table 5: Core follow-up outcome achievement.* U.S. Department of Education. https://nrs.ed.gov/rt/reports/aggregate/2015/all/table-5

National Reporting System for Adult Education. (2016a). *Table 5: Core follow-up outcome achievement.* U.S. Department of Education. https://nrs.ed.gov/index.php/rt/reports/aggregate/2015/all/table-5

National Reporting System for Adult Education. (2016b). *Table 5: Core follow-up outcome achievement.* U. S. Department of Education. https://nrs.ed.gov/rt/tx/2015/table-5

National Reporting System for Adult Education. (2019). *Test benchmarks for NRS educational functioning levels.* U. S. Department of Education. https://nrsweb.org/sites/default/files/NRSTestBnchmrks-082019-508.docx

National Reporting System for Adult Education. (2020a*). Table 1: Participants by entering educational functioning level, ethnicity, and sex.* U. S. Department of Education. https://nrs.ed.gov/rt/reports/aggregate/2019/all/table-1

National Reporting System for Adult Education. (2020b). *Table 1: Participants by entering educational functioning level, ethnicity, and sex.* U. S. Department of Education. https://nrs.ed.gov/rt/tx/2019/table-1

National Reporting System for Adult Education. (2020c). *Table 2: Participants by age, ethnicity, and sex*. U. S. Department of Education. https://nrs.ed.gov/rt/reports/aggregate/2019/all/table-2

National Reporting System for Adult Education. (2020d). *Table 2: Participants by age, ethnicity, and sex.* U. S. Department of Education. https://nrs.ed.gov/rt/tx/2019/table-2

National Reporting System for Adult Education. (2020e). *Table 4a: Educational functioning level gain.* U. S. Department of Education. https://nrs.ed.gov/rt/reports/aggregate/2019/all/table-4a

National Reporting System for Adult Education (2020f). *Table 4a: Educational functioning level gain*. U. S. Department of Education. https://nrs.ed.gov/rt/tx/2019/table-4a

National Reporting System for Adult Education. (2020g). Table 5: Educational functioning level gain. https://nrs.ed.gov/rt/tx/2019/table-5

National Reporting System for Adult Education. (2021). *Technical assistance guide for performance accountability under the Workforce Innovation and Opportunity Act*. U. S. Department of Education. https://nrsweb.org/sites/default/files/NRS-TA-Mar2021-508.pdf

New Mexico Public Education Department. (2017). *2016–2017 ADC manual.* https://webnew.ped.state.nm.us/wp-content/uploads/2017/12/CCRB_graduation_instructinalguides_2016-2017.ADC_.Manual.pdf

Ngo, F., & Kwon, W.W. (2015). Using multiple measures to make math placement decisions: implications for access and success in community colleges. *Research in Higher Education, (56),* 442–470. https://doi.org/10.1007/s11162-014-9352-9

Ngo, F., & Melguizo, T. (2016). How can placement policy improve math remediation? *Educational Evaluation and Policy Analysis, 38*(1). http://doi.org/10.3102/0162373715603504

19 Tex. Admin. Code § 4.51–4.63. (2020). Texas Success Initiative. https://texreg.sos.state.tx.us/public/readtac$ext.TacPage?sl=R&app=9&p_dir=&p_rloc=&p_tloc=&p_ploc=&pg=1&p_tac=&ti=19&pt=1&ch=4&rl=56

19 Tex. Admin. Code §101.4002. (2021). Commissioner's rules concerning substitute assessments for graduation. https://texreg.sos.state.tx.us/public/readtac$ext.TacPage?sl=T&app=2&p_dir=P&p_rloc=186315&p_tloc=&p_ploc=1&pg=7&p_tac=204347&ti=19&pt=2&ch=1&rl=

Noble, J., & Sawyer, R. (2013). *A study of the effectiveness of developmental courses for improving success in college* (ACT Research Report No. 2013-1). ACT. http://www.act.org/content/dam/act/unsecured/documents/ACT_RR2013-1.pdf

Office of Career, Technical, and Adult Education. (n.d.). *College and career readiness (CCR) standards for adult education.* https://sites.ed.gov/octae/2013/04/22/college-and-career-readiness-ccr-standards-for-adult-education

Office of Career, Technical, and Adult Education. (2016*). Implementation Guidelines: Measures and Methods for the National Reporting System for Adult Education*. (ED584415). ERIC. https://eric.ed.gov/?id=ED584415

Office of Career, Technical, and Adult Education. (2021). *Career and Technical Education Title I (Basic grants to states).* https://www2.ed.gov/programs/adultedbasic/funding.html

Office of Vocational and Adult Education. (2013). *An American heritage—Federal adult education: A legislative history 1964–2013*. U.S. Department of Education. https://lincs.ed.gov/publications/pdf/Adult_Ed_History_Report.pdf

Park, T., Woods, C. S., Hu, S., Bertrand Jones, T., & Tandberg, D. (2018). What happens to underprepared first-time-in-college students when developmental education is optional? The case of developmental math and intermediate algebra in the first semester. *The Journal of Higher Education, 89*(3), 318–340. https://doi.org/10.1080/00221546.2017.1390970

Pearson Education. (2013). *THEA Texas higher education assessment: Faculty manual.* http://www.thea.nesinc.com/PDFs/THEA_FacultyManual.pdf

Peirce, C. S. (1905). What pragmatism is. *The Monist 15*(2), 161–181. https://www.jstor.org/stable/27899577

Pennsylvania State Eberly College of Science. (2021). *STAT 509: Design and Analysis of Clinical Trials*. The Pennsylvania State University. https://online.stat.psu.edu/stat509/lesson/18/18.2

Pickard, A. (2016). *WIOA: Implications for low-scoring adult learners.* [ED125478] ERIC. https://eric.ed.gov/?id=EJ1125478

Pimentel, S. (2013). *College and Career Readiness Standards for Adult Education.* MPR Associates. https://lincs.ed.gov/publications/pdf/CCRStandardsAdultEd.pdf

Rowh, M. (2018). Proactive Advising: 5 traits of successful programs. *University Business, 21*(9), 31–34. https://universitybusiness.com/intrusive-advising-for-college-students

Royal, K. D. (2017). Four tenets of modern validity theory for medical education assessment and evaluation. *Advances in Medical Education and Practice, 8*, 567–570. https://doi.org/10.2147/AMEP.S139492

Royer, D., & Gilles, R. (1998). Directed self-placement: An attitude of orientation. *College Composition and Communication*, *50*(1), 54–70. https://doi.org/10.2307/358352

Rutschow, E. Z., Cormier, M. S., Dukes, D., & Cruz Zamora, D. E. (2019). *The changing landscape of developmental education practices: Findings from a national survey and interviews with postsecondary institutions*. Center for the Analysis of Postsecondary Analysis. https://postsecondaryreadiness.org/dev/wp-content/uploads/2019/11/changing-landscape-developmental-education-practices.pdf

Rutschow, E. Z., & Mayer, A. (2018). *Early findings from a national survey of developmental education practices* (Research Brief). Center for the Analysis of Postsecondary Readiness. https://www.mdrc.org/sites/default/files/2018_CAPR_Descriptive_Study.pdf

Rutschow, E. Z., & Schneider, E. (2011). *Unlocking the gate: What we know about developmental education.* MDRC, http://www.postsecondaryresearch.org/i/a/document/18000_unlockingFull.pdf

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance*? Psychological Bulletin 135*, 1–22. http://doi.org/10.1037/a0013978

Schak, O., Metzger, I., Bass, J., McCann, C., & English, J. (2017). *Developmental education: challenges and strategies for reform*. U. S. Department of Education. https://www2.ed.gov/about/offices/list/opepd/education-strategies.pdf

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

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

Scott-Clayton, J., & Rodriguez, O. (2015). Development, discouragement, or diversion? New evidence on the effects of college remediation policy. *Education Finance and Policy 10*(1), 4–45. http://doi.org/10.1162/EDFP_a_00150

Shepard, L. A. (1993). Evaluating test validity. In American Educational Research Association (Ed.), *Review of Research in Education, 19* (pp. 405–450). https://www.jstor.org/stable/1167347

Shields, L. (2016). *Professors' perceptions of the Texas Success Initiative assessment: A phenomenological study* (Publication 10150215) [Doctoral dissertation, Northcentral University]. https://www.proquest.com/docview/1833163089?accountid=5683

Sireci, S. G. (2020). De-"constructing" test validation. *Chinese/English Journal of Educational Measurement and Evaluation, 1*(1), 1–12. https://www.ce-jeme.org/journal/vol1/iss1/3

Stahl, N. A., & Armstrong, S. L. (2018). Re-Claiming, re-inventing, and re-reforming a field: The future of college reading. *Journal of College Reading and Learning, 48*(1), 47–66. http://www.doi.org/10.1080/10790195.2017.1362969

Stout, R. (2014). *The influence of race, ethnicity, sex, time away from secondary education and parental education on college readiness as measured by the Texas Success Initiative Assessment* (Publication 1655431129). [Doctoral dissertation, University of Houston-Clear Lake]. ProQuest Dissertations and Theses Global. https://www.proquest.com/docview/1655431129?accountid=5683

Sunset Advisory Commission. (2013). *Sunset Advisory Commission: Report to the 83rd Legislature.* https://www.sunset.texas.gov/public/uploads/files/reports/TEA%20Staff%20Report.pdf

Sunset Advisory Commission. (2014). *Staff report: Texas Education Agency*. https://www.sunset.texas.gov/public/uploads/files/reports/Report%20to%20the%2083%20Leg%202013.pdf

Talisse, R. B., & Aikin, S. F. (2011). Introduction. In Author (Eds.), *The pragmatism reader: From Peirce through the present* (pp. 9–29). https://www.scribd.com/read/232953749/The-Pragmatism-Reader-From-Peirce-through-the-Present

Tests Determined To Be Suitable for Use in the National Reporting System for Adult Education, 85 Fed. Reg. 47952 (August 7, 2020). https://www.govinfo.gov/content/pkg/FR-2020-08-07/pdf/2020-17301.pdf

Texas Education and Workforce Alignment Act, 87th Leg, R.S., Ch. 2308A, §2308A.003(1), 2021, Tex. Gen. Laws 3767

Texas Higher Education Coordinating Board. (n.d.-a). *Developmental education accountability measures data.* http://www.txhighereddata.org/reports/performance/deved

Texas Higher Education Coordinating Board. (n.d.-b). *Graduation and persistence of developmental education students:2-year institutions (2006–2017)*. https://www.highered.texas.gov/reports/DocFetch.cfm?DocID=1961&Format=XLS

Texas Higher Education Coordinating Board. (n.d.-c). *Graduation and persistence of developmental education students: 4-year institutions (2004–2014)*. https://www.highered.texas.gov/reports/DocFetch.cfm?DocID=1962&Format=XLS

Texas Higher Education Coordinating Board. (n.d.-d). *Texas college and career readiness standards.* http://www.thecb.state.tx.us/institutional-resources-programs/public-universities-health-related-institutions/texas-college-and-career-readiness-standards

Texas Higher Education Coordinating Board. (2008). *Aligning adult basic education and postsecondary education*. http://www.thecb.state.tx.us/DocID/pdf/1628.pdf

Texas Higher Education Coordinating Board. (2014). Transforming Developmental Education in Texas. *Journal of Developmental Education, 38*(1), 32–36. http://www.jstor.org/stable/24614014

Texas Higher Education Coordinating Board. (2015). *THECB updates: TSIA, developmental and adult education.* http://www.thecb.state.tx.us/DocID/PDF/6566.PDF

Texas Higher Education Coordinating Board. (2016). *Developmental education: Updates and progress for underprepared students a report to the Texas Legislature, per Rider 42 and SB 1776, 84th Texas Legislature*. http://www.60x30tx.com/media/1110/developmental-education-updates-and-progress-for-underprepared-students.pdf

Texas Higher Education Coordinating Board. (2017). *Overview: Texas state initiatives*. http://reportcenter.thecb.state.tx.us/reports/data/overview-texas-success-initiative

Texas Higher Education Coordinating Board. (2018a). *Developmental education accountability measures data: Community colleges statewide totals.* http://www.txhighereddata.org/reports/performance/deved/inst.cfm?inst=778877&report_yr=2012&report_type=2

Texas Higher Education Coordinating Board. (2018b). *Developmental education accountability measures data: University statewide totals.* http://www.txhighereddata.org/reports/performance/deved/inst.cfm?inst=445566&report_yr=2018&report_type=4

Texas Higher Education Coordinating Board. (2018c). *Developmental education: Update and 2018-2023 statewide plan for supporting underprepared students: A report to the Texas Legislature per Rider 33, 85th Texas Legislature and SB 1776, 84th Texas Legislature – College Preparatory Courses*. https://reportcenter.highered.texas.gov/reports/legislative/developmental-education-update-and-2018-2023-statewide-plan-for-supporting-underprepared-students

Texas Higher Education Coordinating Board. (2019). *Overview: Texas Higher Education Coordinating Board*. http://reportcenter.thecb.state.tx.us/agency-publication/miscellaneous/overview-texas-higher-education-coordinating-board-2016

Texas Higher Education Coordinating Board. (2020a). *Developmental education accountability measures data: Community colleges statewide totals.* http://www.txhighereddata.org/reports/performance/deved/inst.cfm?inst=778877&report_type=2&report_yr=2020

Texas Higher Education Coordinating Board. (2020b). *Developmental education accountability measures data:* TSTCs/LAMARs *statewide totals.* http://www.txhighereddata.org/reports/performance/deved/inst.cfm?inst=778888&report_type=2&report_yr=2020

Texas Higher Education Coordinating Board. (2020c). *Developmental education accountability measures data: University statewide totals.* http://www.txhighereddata.org/reports/performance/deved/inst.cfm?inst=445566&report_type=4&report_yr=2020

Texas Higher Education Coordinating Board. (2021). *Annual TSI high school summary report student performance on Texas Success Initiative (TSI) readiness measures 2018–2019 high school graduates enrolled in Texas public higher education in fall 2019*. https://reportcenter.highered.texas.gov/reports/data/annual-texas-success-initiative-tsi-report-of-student-performance-statewide-summary-2019-pdf

Texas Workforce Commission. (2017). *Texas adult education and literacy guide.* https://tcall.tamu.edu/docs/17-Texas-AEL-Guide-TWC.pdf

Texas Workforce Commission. (2020). *Texas Adult Education and Literacy assessment guide for program year 2020–2021.* https://www.twc.texas.gov/files/agency/ael-assessment-guide-twc.pdf

Texas Workforce Commission. (2021). *About us*. https://twc.texas.gov/about-texas-workforce

Thayer, H. S. (1966). Pearce and truth: Some reflections. *Transactions of the Charles S. Peirce Society, 32*(1), 1–10. https://www.jstor.org/stable/40320584

Toth, C. (2019). Directed self-placement at two-year colleges: A kairotic moment. *The Journal of Writing Assessment, 12*(1),1–23. https://escholarship.org/uc/item/6g81k736

Truttschel, I. (2010). *An analysis of mathematics TABE Scores as predictors of GED attainment, Postsecondary matriculation and postsecondary degree completion at North Central Technical College in Wausau, WI* (Research Report, University of Wisconsin). http://www2.uwstout.edu/content/lib/thesis/2010/2010truttschelw.pdf

U.S. Department of Education. (2013). *Federal adult education a legislative history 1964–2013*. https://lincs.ed.gov/publications/pdf/Adult_Ed_History_Report.pdf

U.S. Department of Labor. (n.d.). Workforce innovation & opportunity act. https://www.doleta.gov/wioa

Uvin, J. E. (2015). *Program memorandum OCTAE/DAEL 15-4.* https://www2.ed.gov/about/offices/list/ovae/octae-aefla-vision.pdf

Valentine, J. C., Konstantopoulos, S., & Goldrick-Rab, S. (2017). What happens to students placed into developmental education? A meta-analysis of regression discontinuity studies. *Review of Educational Research, 87*, 806–833. https://doi.org/10.3102/0034654317709237

West, C., & Beckman, T. (2018). Concurrent validity. In B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (Vol. 1, pp. 357–357). SAGE Publications. https://www-doi-org.libproxy.txstate.edu/10.4135/9781506326139.n135

Westrick, P. A., & Allen, J. (2014). *Validity evidence for ACT COMPASS placement tests. ACT Research Report Series.* (ED546849). ERIC. https://files.eric.ed.gov/fulltext/ED546849.pdf

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. R., & Schmidt, F. L (2015). College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES. *Educational Assessment, 20*(1), 23–45. http://doi.org/10.1080/10627197.2015.997614

Westrick, P. A., Marini, J. P, Young, L., Ng, H., Shmueli, D., & Shaw, E. J. (2019). *Validity of the SAT for predicting first-year grades and retention to the second year* [Report]. College Board. https://collegereadiness.collegeboard.org/pdf/national-sat-validity-study.pdf

What Works Clearinghouse (2016). *ACT/SAT test preparation and coaching programs*. U.S. Department of Education, Institute of Education Sciences. https://files-eric-ed-gov.libproxy.txstate.edu/fulltext/ED569174.pdf

Whinnery, E., & Pompelia, S. (2019). *Common elements of developmental education policies (*Policy brief). (ED594083). ERIC. http://files.eric.ed.gov/fulltext/ED594083.pdf

Xu, D. (2016). Assistance or Obstacle? The impact of different levels of English developmental education on underprepared students in community colleges. *Educational Researcher 45*(9), 496–507. http://doi.org/10.3102/0013189X16683401

Xu, D., & Dadgar, M. (2018). How effective are community college remedial mathematics courses for students with the lowest mathematics skills? *Community College Review, 46*(1), 62–81. https://doi.org/10.1177/0091552117743789

Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 65–82). Information Age Publishing.

Zumbo, B. D., & Chan, E. K. H. (2014). Reflections on validation practices in the social, behavioral, and health sciences. In B. D. Zumbo & E. K. H. Chan (Eds.), *Validity and validation in social, behavioral, and health sciences* (Vol. 54, pp. 321–327). Springer.

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. *American Educational Research Journal, 42*, 439–464. https://wvde.state.wv.us/abe/documents/CorrelationBetweenVariousPlacementInstruments.pdf