Listening. Learning. Leading.®

Research Report

ETS RR–13-12

Examining Evidence of Reliability, Validity, and Fairness for the *SuccessNavigator*™ Assessment

Ross Markle

Margarita Olivera-Aguilar

Teresa Jackson

Richard Noeth

Steven Robbins

July 2013

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

Examining Evidence of Reliability, Validity, and Fairness for the *SuccessNavigator*™ Assessment

Ross Markle, Margarita Olivera-Aguilar, and Teresa Jackson
Educational Testing Service, Princeton, New Jersey

Richard Noeth
Independent Contractor, Iowa City, Iowa

Steven Robbins
Educational Testing Service, Princeton, New Jersey

July 2013

Action Editor: Brent Bridgeman

Reviewers: Richard J. Tannenbaum and Jeremy Burrus

Abstract

The *SuccessNavigator*™ assessment is an online, 30 minute self-assessment of psychosocial and study skills designed for students entering postsecondary education. In addition to providing feedback in areas such as classroom and study behaviors, commitment to educational goals, management of academic stress, and connection to social resources, it is also designed to predict a range of early academic outcomes. By indicating students' likely success, advisors, faculty, and staff can target their interactions with students to increase their likelihood of success. This report outlines evidence of reliability, validity, and fairness to demonstrate the appropriateness of SuccessNavigator for these purposes.

Key words: psychosocial factors, noncognitive factors, college student retention, postsecondary education, student success

Table of Contents

List of Tables

List of Figures

Overview

Psychosocial (also referred to as *noncognitive*) factors refer to personality, motivation, study skills, intrapersonal and interpersonal skills, and a wide array of factors outside of cognitive ability. Research over the past decade has provided extensive evidence showing the relevance of psychosocial skills to success in higher education (see Porchea, Allen, Robbins, & Phelps, 2010; Poropat, 2009; Robbins, Allen, Casillas, Peterson, & Le, 2006; Robbins, Lauver, Le, Davis, Langley, & Carlstrom, 2004; Trapmann, Hell, Hirn, & Schuler, 2007). These studies have shown that psychosocial skills have a strong relationship to academic outcomes such as grades and enrollment persistence and that this relationship is significant, even when controlling for academic (e.g., standardized admissions or placement test scores and high-school grade point average, or GPA) and socioeconomic factors. Accordingly, the *SuccessNavigator*™ assessment was developed to provide a quality assessment of psychosocial skills to students, faculty, staff, and administrators in higher education. This assessment is designed to serve four general purposes:

1. To identify entering students' overall likelihood of enrollment success (persistence to the second year of college) and academic success (first-year college GPA).
2. To provide feedback on academically related strengths or hurdles that students may face, both directly to students and those that work with them. This feedback includes co-curricular supports that could be used to increase students' likelihood of success.
3. To provide information above and beyond traditional placement or admission tests that could be used to make more accurate course placement decisions, specifically to guide the acceleration of students into higher levels of mathematics and English courses.
4. To aggregate student scores within an institution, such that administrators can make decisions about programs and services that could better meet the needs of incoming students.

Stakes and Consequences of SuccessNavigator

Many have argued the intended uses and consequences of an assessment are important parts of validity (e.g., Kane, 2013; Messick, 1994; 1995; Moss, 1992) and should be appropriately considered by test developers (Reckase, 1998). Generally, assessments are referred to as *low-stakes* when there are no consequences to the individual, such as when students are assessed for institutional evaluation or program improvement. However, even though SuccessNavigator refers to the individual test-taker, SuccessNavigator is considered a lower stakes assessment, given its intended uses. Mehrens (1998) cited an example of low-stakes assessment as one used for "planning specific classroom interventions for individual students" (p. 4), which greatly resembles much of the intended use of SuccessNavigator. In addition, high-stakes assessments generally lead to positive consequences for those who score well and negative consequences for those who don't (e.g., admission vs. rejection to college). With SuccessNavigator, the primary consequence for students with lower scores is that they might receive more resources to support their success. Moreover, we firmly discourage the use of SuccessNavigator for traditional high-stakes situations such as admissions.

To be sure, providing information that could improve a student's chance to persist and receive a degree is a significant outcome for the test taker, yet the setting, consequences, and intended use of this assessment do not match traditional criteria for high-stakes tests. Ultimately, this designation as a lower-stakes assessment plays a key role in the demonstration of evidence and the intended use of the assessment.

Purpose of Current Report

This report is designed to demonstrate the psychometric quality of SuccessNavigator using well established validity frameworks (Kane, 2006; Messick, 1995) and more specifically using a method and terminology demonstrated by Benson (1998). Under these frameworks, many common aspects of validity evidence (e.g., content validity, predictive validity) all fall under the auspices of *construct validity*.

In the sections that follow, we first discuss the relationship between existing theory and research and the scale and item development processes, which provide evidence of substantive (or *content*) validity. This type of evidence demonstrates that a given assessment adequately and accurately represents the theoretical framework upon which it is based.

Second, we discuss the factor analytic procedures that examined the internal relationships among SuccessNavigator's items and scales, which provide evidence of *structural* validity. This type of evidence shows that the variables within an assessment (i.e., items and scales) relate to each other as hypothesized.

Third, we discuss the development of SuccessNavigator's success indices, which predict the academic outcomes of college GPA, persistence, and course grades in mathematics and English. Such relationships with variables outside of an assessment, to the extent that they align with hypotheses, provide evidence of *external* validity. As part of external validity, we also present issues related to fairness, because evidence of test bias, such as differential measurement or prediction across gender or race/ethnicity, can be a significant threat to the validity of inferences drawn from an assessment.

Substantive Validity

SuccessNavigator uses a hierarchical framework that includes four broad areas, referred to as *general skills*. These four areas are academic skills, commitment, self-management, and social support (see the appendix). Each general skill contains more granular facets, or subskills. These general skills and subskills were developed using an approach that blended theory, practice, and data. The sections below provide evidence of how these constructs were supported and used in item development.

Alignment to Theory

From a theoretical perspective, the SuccessNavigator general skills were based, in part in the Big 5 theory of personality (see McCrae & Costa, 1987). One personality factor—conscientiousness, referring to one's organization, timeliness, effort, and drive to achieve goals—has been shown to have particular relevance to student success in higher education (Poropat, 2009). As such, two general skills relate to the behavioral (academic skills) and goal-related (commitment) components of conscientiousness.

Academic skills refer to a student's use of behaviors and strategies to manage time and work (organization) and tendency to attend class, participate effectively, and complete assignments in a timely manner (meeting class expectations). Commitment contains two subskills: commitment to college goals, which refers to intentionality toward and perceived value of a college degree, and institutional commitment, which refers to the attachment toward one's institution. The organizational and task-related behaviors that are part of academic skills, as well as the goal-related intentionality that is part of commitment have both been empirically demonstrated as facets of conscientiousness (MacCann, Duckworth, & Roberts, 2009).

Other studies have shown areas of emotional stability (also referred to by its opposite, neuroticism) to be relevant to student success (see Poropat, 2009; Richardson, Abraham, & Bond, 2012; Robbins et al., 2004; Robbins et al., 2006; Trapmann et al., 2007). Emotional stability refers to one's tendency to avoid undue stress or worry (McCrae & Costa, 1987). The general skill of self-management directly addresses this factor through the subskills of sensitivity to stress (general susceptibility to stress in academic situations) and test anxiety (thoughts and physical reactions to test situations).

In our construct map, academic self-efficacy (perception of one's own ability to succeed in college) is also included under self-management. Although some models align self-efficacy with conscientiousness (e.g., Judge, Erez, Bono, & Thoresen, 2002), we have included it here because of its tie to adjustment and other self-regulatory processes (e.g., Bong & Skaalvik, 2003; Chemers, Hu, & Garcia, 2001.

Finally, the tendency to connect with others has also been shown to relate to academic success (Karabenick, 2003; Napoli & Wortman, 1996; Robbins et al., 2004, Robbins et al., 2006; Trapmann et al., 2007). Thus, the general skill of social support was developed and includes three subskills: connectedness, institutional support, and barriers to success. Connectedness refers to one's general relationship to others. Institutional support measures students' tendency to engage with formal resources on campus. Lastly, barriers to success refer to factors outside of academic life, such as financial constraints and family obligations that might impede or foster a student's success. Interpersonal connectedness relates to the personality domain of extraversion, again showing similarity between the Big 5 model and SuccessNavigator.

Admittedly, research has not shown each personality factor to be equally predictive of success. Generally, conscientiousness has been supported as the most empirically relevant factor toward predicting achievement in higher education, even equaling the effect of cognitive ability (Poropat, 2009). Moreover, Robbins et al. (2009) found that academic and class-related factors were far more predictive of success than social or emotional factors. However, our intent in developing SuccessNavigator was not merely to predict success, but to provide a holistic understanding of each student. Although certain factors may demonstrate lesser predictive efficacy when considered across the population as a whole, complex mediated and moderated relationships often underlie student success. By including a wide range of factors, faculty and staff can better understand and work with each student and seek to understand individual paths to success.

Alignment to Practice

In addition to established theories of personality, the development of SuccessNavigator was informed by practice in two important ways. First, given that SuccessNavigator is designed to connect students with resources on college campuses, the general skills and subskills were also designed to align with many of the success-related programs and services that are offered by colleges and universities.

Many efforts on campus work to improve academic skills by improving organizational skills, study skills, or classroom behavior (Robbins et al., 2009). Such efforts may be housed under student success or first-year experience courses, student success courses, or tutoring efforts (Kuh, Kinzie, Buckley, Bridges, & Hayek, 2006).

O'Banion (2013) argued that effective advising programs should seek to understand students' life and career goals in order to effectively align degree aspirations, institutional selection, program enrollment, and course selection. Students cannot be fully driven to attend an institution and achieve a degree until they fully understand how it fits in their longer term goals. As such, commitment and its subskills align well with academic advising, career counseling, and other goal-related efforts.

According to the *Standards and Guidelines for Counseling Services* (Council for the Advancement of Standards in Higher Education [CAS], 2011), counseling services at institutions of higher education are charged with addressing many emotional and psychological issues. These range from academic-related stressors to personal problems to considerable issues of mental health. SuccessNavigator's self-management general skill addresses several factors related to academic life and can be used by students, faculty, and staff to connect students to counseling resources that can help them manage these stressors.

Additionally, many programs and services exist to facilitate students' integration into the social aspects of college life. These programs relate to a long-standing finding in higher education that students who are more involved in activities on campus are more likely to do well in class and persist to a degree (cf. Kuh et al., 2006). These include orientation, student life, and first-year experience programs. However, some research has shown that social integration is important as well for nontraditional students, such as commuters and students not entering directly out of high school (Sorey & Duggan, 2008). As such, the general skill of social support addresses social connections that can be integral to the success of many students.

In addition to alignment with programs and services on campus, SuccessNavigator was also informed by practitioners themselves. Faculty, staff, administrators, and students from nearly 50 higher education institutions reviewed and confirmed the relevance of the SuccessNavigator construct map (general skills and subskills) to student success.

Overall, the alignment of these general skills and subskills to existing institutional expectations and practices not only provides validity evidence, but may foster adoption of the assessment by relating the relevant constructs to the existing practices of score users. Moreover, this not only helps students and advisors to make inferences about the student, but can guide connections to on-campus resources to foster those skills, if needed.

Lastly, the structure of SuccessNavigator scores was determined by empirical study. By analyzing the reliability and factor structure of test scores within our proposed test framework, we were able to determine which scales performed well and were to be retained, which scales performed poorly and were to be removed, and which scales required modification. This process is described further in the Structural Validity section of this report.

Item Development

In developing the items for SuccessNavigator, a team of item writers from within ETS was assembled, consisting of five doctoral-level researchers, one each specializing in social psychology, counseling psychology, and assessment and measurement psychology and two specializing in educational psychology. Each researcher had extensive experience in both psychosocial assessment and college student success.

In group meetings over the course of one week, the team met first to discuss the construct map discussed above and presented in the appendix. The framework of four general skills and 10 subskills was first reviewed by the item writing team, which made minor changes to several operational definitions in order to focus the item-writing process. For each subskill, the team was instructed to develop 10–15 self-report, Likert-type statements with an overall goal of a 100–150 item assessment, which would take roughly 30 minutes for a test taker to complete. The construct map and these parameters for assessment length served as the test blueprint for the item writing team.

Items were developed collaboratively. To begin the exercise, one of the subskills would be selected, and team members would suggest potential items to address that factor. The team would then consider the item for adherence to the operational definition, appropriate language (e.g., reading level, colloquialisms, subgroup bias), and potential sources of random or systematic error in responses.

Ultimately, 125 self-report items with a 6-point (strongly disagree to strongly agree) response scale were developed. More items than were required by the test specifications were developed with the understanding that certain items would be removed during pilot testing. For some subskills, the team identified and adapted items from previous ETS research projects (e.g., Liu, Rijmen, MacCann, & Roberts, 2009; Roberts, Schulze, & MacCann, 2007) that could be applied to the construct map. These preexisting items comprised roughly 15-20% (20-30 items) of the initial item pool.

Once the initial item pool was developed, each item went through an internal fairness review that considered if items were appropriate for all potential users and subgroups (e.g., race/ethnicity, English language learners). No items were flagged for issues of fairness. The next step was to administer these items in our national field trial to determine the psychometric quality of the items and scales, removing items that did not adequate address the constructs of interest or were redundant in doing so. The process for refining the initial item pool down to the final assessment is described in the section that follows.

Structural Validity

Empirical Determination of Scales and Dimensionality

In order to test the psychometric properties of the initial 125 items, they were administered to a large, multi-institutional sample of entering college students. The items were randomly presented to each student using an online survey platform. All negatively keyed items were reverse-coded to ensure that high values in all items indicated a desirable level in the subskill being evaluated.

A unidimensional model was fit to the items of each subskill using confirmatory factor analysis (CFA). This analysis was conducted with the purpose of selecting items with adequate psychometric properties while also gathering evidence of the structural validity of SuccessNavigator. Once the items with adequate statistical requirements were selected, the fit of the unidimensional model was examined. Good fit indicated that the items were relating to one another as hypothesized and was regarded as evidence of structural validity.

The reasoning behind this approach (i.e., treating each scale as a separate, unidimensional model, rather than testing a hierarchical structure) was that, although it is believed that each subskill represents a latent variable that explains the answers to the items in that construct, it is not the case that each skill is conceptualized as a latent variable explaining the subskills. That is, each skill combines two or three subskills to form a construct, without any assumptions about the relationships between them. For example, it is not the case that there is a social support latent variable that causes the response to the subskills of connectedness, institutional support and instrumental support. Instead, the combination of these subskills forms the construct of social support. In this sense, it was relevant to examine each subskill as unidimensional, but no hierarchical relationships were tested in the factor analysis.

Sample. Data were gathered from nine 2- and 4-year institutions from across the United States during the summer and fall of 2012. Participating institutions were recruited from various ETS advisory boards, including the National Community College Advisory Council (2-year institutions) and Higher Education Advisory Council (4-year institutions). These groups are designed to represent an array of institutional characteristics (e.g., size, urban vs. rural setting, geography), which increased the diversity of participants. Ultimately, all nine institutions were public colleges and universities, with five 4-year schools and four 2-year institutions. Overall, the institutions were diverse in terms of their geography, with institutions in the western, southern, eastern, and midwestern sections of the United States. Enrollments ranged from roughly 5,000 students to more than 20,000. Nearly 60% of the sample came from two participating institutions (one 2-year and one 4-year). This variance in sample size arose from the fact that schools were allowed to design their recruitment and administration procedures, which lead to some schools seeking larger samples than others.

Table 1 shows the total sample sizes per institution and by gender and race. The sample consisted of a total of 5,618 students, from which 5,120 students (91%) had complete data (i.e., responses to all items) and 5,061 students provided demographic information. Only students with complete data were included in the analyses.

The sample was composed of slightly more females (n = 2,983 ; 58.9%), and was diverse in terms of race/ethnicity: 38% Hispanic/Latino (n = 1,923), 37.8% White (n = 1,911), 8.1% African American (n = 408), 218 Asian (4.3%), 70 Native American (1.4%), 22 Pacific Islander (0.2%), 303 multiracial/multiethnic (5.9%), 106 (2.1%) indicated their ethnicity as other, and 111 (2.2%) preferred not to answer.

Table 1

Sample Used in Selecting Items and Model Fit Evaluation by Institution, Gender, and Race/Ethnicity

Institution	*N*	Gender		Race/ethnicity						
		Male	Female	White	African American	Native American	Hispanic/ Latino	Asian	Pacific islander	Multi-racial
1	1,781	815	958	209	207	3	1,031	85	7	129
2	1,182	423	750	370	22	63	569	39	1	73
3	549	275	255	265	58	1	120	27	2	25
4	456	135	319	355	18	0	47	10	0	18
5	404	182	222	310	34	1	34	2	0	13
6	225	57	167	151	25	1	16	9	11	17
7	207	64	125	107	29	1	22	7	1	11
8	172	62	108	47	11	0	52	38	0	12
9	144	65	79	97	4	0	32	1	0	5
Total	5,120	2,078	2,983	1,911	408	70	1,923	218	22	303

Analysis. CFAs were used to fit unidimensional models to each subskill using the statistical software M*plus* 6.1 (Muthén & Muthén, 1998–2010). The results of the CFAs were used to evaluate the psychometric properties of the items—namely standardized loadings and communalities—and to judge the fit of the unidimensional model. The items that showed standardized loadings larger than 0.2 and communality values larger than 0.1 were considered to be representative of the construct measured by the general skill. Items with lower standardized loadings and communalities were kept only if they were considered to be crucial to measure the subskills and if the overall fit of the model remained adequate with their inclusion.

After all of the items with adequate psychometric properties were selected, the construct validity of each scale was evaluated using an internal approach. Because each scale was conceptualized as measuring one construct, a unidimensional model with acceptable fit to the data was regarded as evidence of structural validity.

When examining model fit within each subskill, correlations between unique variances were allowed when the items were related beyond what would be expected by a unidimensional model. For example, negatively worded items were allowed to have correlated unique variances, and items with similar wording were allowed to correlate when necessary. One instance of similar wording included academic self-efficacy items with the clause “If I apply myself,” which was included so that students clearly understood that effort was implied.

The overall fit of the model was examined using the chi-square of model fit. Because the chi-square fit statistic is sensitive to small deviations from the model when large sample sizes are used, three other fit indices were also considered: the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). A model was considered to have an adequate fit to the data when the values of the fit indices were CFI > .90, RMSEA and SRMR < .08; while models with values of CFI > .95, RMSEA and SRMR < .05 were considered to have a good fit to the data (Browne & Cudeck, 1993; West, Taylor, & Wu, 2012).

In order to diminish the risk of sample-specific conclusions, the total sample was randomly divided into two roughly equal samples. The sizes for Samples 1 and 2 were 2,529 and 2,591, respectively. Sample 1 was used to develop a model with optimum fit by iteratively identifying items with low standardized loadings, removing them from the model, and retesting model fit. Once a model with an adequate fit to the data was achieved, the final model was fit to

Sample 2. The purpose of conducting the analyses in two different samples was to ensure that the conclusions were not dependent of the characteristics of the sample at hand and that the results could be generalized to future samples. The final selection of items was based on the results in Sample 1.

Results. The final selection of items for each subskill, including the item loadings and communalities (R^2), can be found in Table 2. Only the organization and the barriers to success subskills have items with loadings or communalities below the cutoffs of .2 and .1 respectively. These items were retained because they were considered important representations of the underlying theory. Overall, 93 of the 125 items were retained (74%). The loss of items for each subscale varied, with some scales retaining all their original items and others losing as many as six items.

Table 2

Item Loadings and Communalities (R^2) for SuccessNavigator Subskills

Subskill	Item	Item	Loading	R^2
Meeting class expectations	BC1	1	.660	.436
	BC2	2	.699	.489
	BC3	3	.562	.316
	BC4	4	.625	.391
	BC5	5	.678	.459
	BC7	7	.386	.149
	BC8	8	.485	.236
	BC9	9	.328	.108
	BC10	10	.517	.267
	BC12	12	.628	.394
Organization	Org2	2	.488	.238
	Org3	3	.634	.402
	Org6	6	.685	.469
	Org8	8	.641	.411
	Org5r	5[a]	.162	.026
	Org7r	7[a]	.327	.107
	Org11	11	.868	.753
	Org10r	10[a]	.200	.040
	Org14r	14[a]	.371	.137
Commitment to college goal	DG1	1	.762	.581
	DG2	2	.770	.593
	DG3	3	.808	.653
	DG4	4	.709	.503
	DG5	5	.454	.206

Subskill	Item	Item	Loading	R^2
	DG7	7	.332	.110
	DG10	10	.700	.490
	DG11	11	.591	.349
	DG12	12	.742	.551
Institutional commitment	InC2	2	.518	.258
	InC7	7	.804	.646
	InC8	8	.769	.591
	InC9	9	.781	.610
	InC10	10	.832	.693
	InC11	11	.604	.365
	InC14	14	.765	.585
	InC15	15	.811	.657
Sensitivity to stress	STRS1r	1[a]	.439	.193
	STRS3	3	.779	.607
	STRS10r	10[a]	.535	.287
	STRS11	11	.613	.376
	STRS12r	12[a]	.454	.206
	STRS2r	2[a]	.509	.259
	STRS4r	4[a]	.550	.302
	STRS7r	7[a]	.510	.261
	STRS9r	9[a]	.569	.324
	STRS15	15	.692	.479
Test anxiety	TDUR10r	1[a]	.758	.575
	TDUR4r	4[a]	.735	.541
	TPRE4r	4[a]	.680	.462
	TDUR5r	5[a]	.667	.444
	TPST5r	5[a]	.737	.543
	TPRE5r	5[a]	.404	.163
	TDUR6r	6[a]	.419	.176
	TDUR7r	7[a]	.774	.599
	TDUR8r	8[a]	.748	.560
Academic self-efficacy	ASE2	2	.592	.350
	ASE3	3	.635	.403
	ASE4	4	.590	.350
	ASE5	5	.761	.578
	ASE6r	6[a]	.551	.304
	ASE8	8	.742	.551
	ASE9	9	.622	.387
	ASE11r	11[a]	.597	.356
	ASE12r	12[a]	.574	.330
Connectedness	CNCT1	1	.620	.384
	CNCT2	2	.721	.520
	CNCT4	4	.799	.638

Subskill	Item	Item	Loading	R^2
Connectedness	CNCT5	5	.765	.585
	CNCT7	7	.757	.573
	CNCT8	8	.681	.464
	CNCT10	10	.483	.233
Institutional support	NSUP1	1	.493	.243
	NSUP2	2	.706	.498
	NSUP3	3	.421	.178
	NSUP4	4	.730	.533
	NSUP5r	5[a]	.442	.195
	NSUP6r	6[a]	.571	.326
	NSUP9	9	.554	.307
	NSUP10	10	.711	.506
	NSUP11	11	.523	.274
	NSUP12r	12[a]	.587	.344
	NSUP13r	13[a]	.558	.312
Barriers to success	Strm1r	1[a]	.347	.120
	Strm2r	2[a]	.324	.105
	Strm3r	3[a]	.164	.027
	Strm4r	4[a]	.215	.046
	Strm5	5	.817	.668
	Strm6	6	.586	.344
	Strm7	7	.517	.267
	Strm8r	8[a]	.319	.102
	Strm9r	9[a]	.172	.030
	Strm11	11	.785	.617
	Strm10r	10[a]	.297	.088

[a] Reverse-scored item.

The overall fit indices of the unidimensional model for each subskill in the two samples are shown in Table 3. According to the CFI and the SRMR, a unidimensional model was found to have a good fit to the data in every subskill. RMSEA values less than 0.05 were observed in six subskills, which suggest that a unidimensional model shows good fit in meeting class expectations, organization, sensitivity to stress, academic self-efficacy, institutional support, and barriers to success. The rest of the subskills had RMSEA values less than 0.08, indicating that a unidimensional model has an adequate fit to the data in commitment to college, institutional commitment, text anxiety, and connectedness.

Table 3

Fit Indices for Subskill-Level Confirmatory Factor Analysis (CFA) Models

Subskill	Sample 1						Sample 2				
	Items	X^2	*df*	CFI	RMSEA	SRMR	X^2	*df*	CFI	RMSEA	SRMR
Meeting class expectations	10	164.30	29	.984	.043 (.037, .049)	.027	194.90	29	.981	.047 (.041, .053)	.026
Organization	10	42.76	15	.996	.027 (.018, .037)	.015	83.54	15	.990	.042 (.033, .051)	.019
Commitment to college	7	436.72	27	.956	.077 (.071, .084)	.035	358.40	27	.966	.069 (.063, .075)	.034
Institutional commitment	9	184.12	19	.985	.059 (.051, .066)	.019	120.12	19	.991	.045 (.038, .053)	.020
Sensitivity to stress	13	73.16	14	.995	.041 (.032,.050)	.013	55.65	14	.996	.034 (.025, .043)	.010
Test anxiety	9	242.52	27	.976	.056 (.05, .063)	.023	250.31	27	.977	.056 (.05, .063)	.023
Academic self-efficacy	10	90.09	21	.992	.036 (.029, .044)	.015	102.57	21	.991	.039 (.031, .046)	.017
Connectedness	7	94.09	14	.989	.048 (.039, .057)	.017	135.86	14	.983	.058 (.049, .067)	.019
Institutional support	13	234.28	36	.980	.047 (.041, .052)	.027	177.86	36	.987	.039 (.033, .045)	.023
Barriers to success	11	111.11	23	.988	.039 (.032, .046)	.023	135.62	23	.985	.043 (.037, .051)	.024

Note. CFI = comparative fit index, RMSEA = root mean square error approximation, SRMR = standardized root mean square residual.

Scale Intercorrelations

While the CFAs provided structural evidence to support scores at the subskill level, convergent and discriminant correlations (see Table 4) were used to support the structuring of those scores within the larger general skills. Convergent and discriminant validity refer to the extent that scales are correlated or uncorrelated, respectively, in alignment with hypotheses. For example, two scales that are hypothesized to be similar and have a high observed correlation provide evidence of convergent validity. Scales that are hypothesized to be unrelated and subsequently are found to be so provide evidence of discriminant validity.

In terms of convergent validity, it was expected that the subskills would present moderate correlations (bivariate *r*'s between .3 and .5) within each general skill. As shown in Table 4, each of the subskills was moderately correlated (ranging from .344 to .556) with other scales within the general skill. The one exception is the low correlation between academic self-efficacy and test anxiety (r = .185; see Table 5). Because of this finding, academic self-efficacy was not included in the general skill score for self-management, though it is reported with the test anxiety and sensitivity to stress due to theoretical similarities.

With regard to discriminant validity, it was expected that the subskill scores from different general skill areas would show small ($r \approx .10$) to medium correlations ($r \approx .30$). Table 4 shows that many (22 of 37) of the correlations across subskills from different general skills do not exceed an absolute value of .4. However, in some cases, high correlations among subskills from different general skill areas were expected. For instance, given that academic self-efficacy has been related to other aspects of conscientiousness, the high correlation between it and meeting class expectations (r = .62) was not surprising. Ultimately, the results presented in Table 4 were generally supportive of the hypothesized construct map. Correlations within the general skill areas were moderate, with all but two values near or above .5 (suggesting agreement), but never exceeding .6 (suggesting unique contributions from each scale).

Reliability

The reliability for each of the subskills, using the final set of items, was obtained through Cronbach's alpha. Table 6 shows the reliability for each subskill as well as other descriptive information. All scales had reliability values above the suggested standard of .70 that would be applied to a low-stakes self-report assessment such as SuccessNavigator (Nunnally, 1978).

Table 4

Correlations Between the SuccessNavigator Subskills

General skill/subskill	1.	2.	3.	4.	5.	6.	7.	8.	9.
Academic skills									
1. Meeting class expectations									
2. Organization	**.533** [b]								
Commitment									
3. Commitment to college	.493	.263							
4. Institutional commitment	.423	.240	**.540** [b]						
Self-management									
5. Stress sensitivity[a]	.364	.137	.196	.208					
6. Test anxiety[a]	.123	.030	-.058	.002	**.514** [b]				
7. Academic self-efficacy	.619	.284	.588	.424	**.469** [b]	**.185** [b]			
Social support									
8. Connectedness	.494	.291	.344	.468	.380	.176	.357		
9. Institutional support	.614	.381	.386	.393	.444	.221	.512	**.556** [b]	
10. Barriers to success[a]	.437	.250	.393	.320	.387	.195	.596	**.344** [b]	**.501** [b]

[a] Scales are reverse-coded so that higher scores indicate positive approaches to success (e.g., high “test anxiety” indicates less anxiety during test situations).

[b] Bolded correlations signify those between subskills within the same general skill.

Table 5

Correlations Between the SuccessNavigator General Skills

General skill	1	2	3
1. Academic skills			
2. Commitment	.446		
3. Self-management	.206	.140	
4. Social support	.557	.540	.411

Note. All correlations significant at $p < .001$.

Table 6

Reliability and Descriptive Statistics for SuccessNavigator Subskills

General skill/subskill	Item	Alpha	*M*	SD	Skewness	Kurtosis
Meeting class expectations	10	.83	4.87	.59	-.47	.94
Organization	9	.80	3.91	.08	-.07	-.27
Commitment to college	9	.84	5.34	.56	-1.58	4.77
Institutional commitment	8	.90	5.06	.70	-1.07	2.31
Sensitivity to stress	10	.88	4.12	.87	-.30	0
Test anxiety	9	.88	3.28	1.04	.30	-.21
Academic self-efficacy	9	.86	5.20	.62	-.90	.84
Connectedness	7	.86	4.26	.86	-.60	.60
Institutional support	11	.86	4.57	.71	-.37	-.02
Barriers to success	11	.78	4.64	.68	-.47	.16
Academic skills	19	.86	4.44	.60	.05	-.30
Commitment	17	.89	5.25	.48	-.56	-.07
Self-management	19	.89	3.79	.77	.30	-.34
Social support	29	.89	4.57	.56	-.09	-.26

In addition, reliabilities for the general skill scales are also provided and also exceeded the criterion for sufficiency. All but one general skill exceeded .9, with the lone exception, academic skills, approaching that value (α = .86). Admittedly, these scales also have more items, which inherently leads to higher score reliability.

Additionally, Kline (2010) recommended absolute values of 3 and 10 as criteria for normality when examining skewness and kurtosis, respectively. Given that none of the subskills exceed those values, each appears to demonstrate reasonable univariate normality.

SuccessNavigator Success Indices

SuccessNavigator reports four compound indices to advisors, each designed to predict a specific academic outcome. The indices optimally weight high school GPA, standardized test scores, and SuccessNavigator subskills scores to maximally predict success. Each of these components addresses a key student characteristic that contributes to success. In a previous section, we reviewed some of the research that supports the relationship between SuccessNavigator scores and student success. However, the more traditional variables of high school GPA and standardized test scores are still important to include.

Standardized tests used for placement and admissions, such as SAT®, ACT, Accuplacer®, and COMPASS, are well-established, and perhaps the most widely used indicators of student success (see Camara & Echternacht, 2000; Mattern & Packman, 2009; Mattern & Patterson, 2009; Robbins et al., 2004). Certainly, cognitive ability has a role in student success and is an important factor to include in a holistic model. Research supporting the relevance of psychosocial skills has not intended to show cognitive ability to be an ineffective predictor of student success, merely an incomprehensive one.

Indeed, as Wiley, Wyatt, and Camara (2010) noted, models that include multiple indicators of student success prove more effective than those that include any single indicator. They, and many others, promoted the use of high school GPA as an additional predictor of success. High school GPA is a compelling variable because it includes some indication of academic ability, but also psychosocial factors such as class participation, timeliness, and interpersonal skills (Sawyer, 2010).

However, the inherent complexity of high school GPA also creates concerns. Given that teachers may vary in their use of academic performance, classroom behavior, and other factors in determining grades, high school GPAs may not be comparable across students. Several have noted that these varying practices decrease the reliability of grades (e.g., Allen, 2005; Brookhart, 1993; Burke, 2006), thus limiting the efficacy of high school GPA as a sole predictor. Another factor is the varying comparability of grades across high schools, due to differences in academic rigor (see Adelman, 1999, 2006).

Ultimately, the importance of a holistic model of student success is rooted in the understanding that any one indicator has positive and negative features. The three potential inputs to the SuccessNavigator success indices—test scores, high school GPA, and psychosocial skills—all relate to success, but each is insufficient on its own. Moreover, as Robbins et al. (2004) noted, the importance of these factors vary depending upon the outcome being studied.

The academic success index is designed to predict students' overall college GPA, indicating a student's likelihood of performing well in the classroom. Previous meta-analytic findings have shown that high school GPA and test scores are strong predictors of college GPA, though psychosocial factors also contribute significantly to the model (Robbins et al., 2004).

The retention success index indicates a student's likelihood to persist to a second semester using institutional retention data. We define retention here according to traditional dichotomous definitions, with 1 indicating a return to the institution and 0 indicating departure (including students who transfer). The Robbins et al. (2004) meta-analysis actually showed that psychosocial factors more strongly relate to persistence than high school GPA or standardized test scores.

Finally, the math and English course acceleration indices are tied to a student's likelihood to succeed in college-level courses, using course-level grades as the outcome. Comprehensive meta-analyses (i.e., including high school GPA, tests scores, and broad-based psychosocial data) predicting course grades have not been conducted. Thus, there are no hypotheses about the relative weights of academic and psychosocial predictors, either for comparing models for course grade outcomes to those used for predicting college GPA or persistence or for comparing a model predicting success in mathematics to one predicting success in English.

For the academic and retention success indices, scores are designed to generally indicate the likelihood of success to faculty, staff, or advisors so that they can best determine the level of engagement that a student might require in order to be successful. As such, scores are broken into low, medium, and high score bands in order to simplify interpretation. It should be noted that the success indices are not reported to students directly, given that research into score feedback suggests that negative feedback (i.e., informing students that they have a low probability of success) may have a negative impact on their success (e.g., Bridgeman, 1974).

The course acceleration indices are designed with a more specific function. Given recent research showing high rates of failure for students placed into developmental courses, several studies have suggested that assessments of psychosocial skills should be included in course placement decisions, given that placement tests measure cognitive ability and do not indicate all of the factors that might determine a student's success (Boylan, 2009; Conley, 2007). Thus, the course acceleration indices are designed to be used in concert with existing placement tests to make recommendations for students who could be accelerated to higher levels of mathematics or English.

That is, if a student places into a developmental mathematics course, but an advisor feels that the student's holistic set of scores—including placement tests and SuccessNavigator scores—suggest the student could succeed in a higher level course (which could still be in the developmental sequence), that student could be accelerated. However, given compelling evidence of low rates of success in developmental courses and the effectiveness of accelerating students into higher levels (Complete College America, 2012; Scott-Clayton, 2012), students who indicate a lower likelihood of success are also recommended for acceleration, but with caution, meaning that they will likely require additional support (e.g., tutoring, advising, supplemental instruction) to succeed in their courses.

Given this intended use, the final analyses for the course acceleration indices exclude test scores because of the many levels of courses at which these acceleration decisions are made. If the indices were to include placement tests or similar indicators of cognitive ability, students at lower levels of ability—who would thus be recommended for lower levels of mathematics and English courses—would inherently be less likely to be recommended for acceleration. Again, rather than including these test scores in the model of student success, the course acceleration indices are designed to be used in concert with placement tests.

In developing these four indices—academic success, retention success, math course acceleration, and English course acceleration—a series of regression analyses were conducted in which the outcome variables—first-semester college GPA, first-semester persistence, grades in college mathematics courses, grades in college English courses, respectively—were predicted from the 10 SuccessNavigator subskills and high school GPA. With academic success and retention success, standardized placement and admissions test scores were also included in the model. As mentioned, test scores were excluded from the models supporting the course acceleration indices.

Variables

All outcome data (i.e., college GPA, retention, and course grades in English and mathematics) were provided from institutional records and were not self-reported by students. First semester college GPA was scored on a 0–4 scale, with values of 0 included as valid scores. Persistence was scored dichotomously, referring to whether or not the student returned for the following semester to the same institution. Finally, course grades for college-level mathematics and English were scored on a 0–4.0 scale (F, withdraw, or incomplete = 0; D = 1; C = 2; B = 3; A = 4).

At the time of the present study, only first-semester outcomes were available. Thus, for the purposes of this study, first semester persistence and college GPA served as proxies for success measures that would, ideally, be measured over the first full year. Future research will include more data as they becomes available.

All 12 potential predictors (10 SuccessNavigator subskills, high school GPA, and test scores) were converted into standardized scales with a mean of 100 and a standard deviation of 15. This is done as part of the scoring of SuccessNavigator in order to facilitate interpretation of scores. By placing all factors on the same scale, users can more easily interpret differences (whereas differences on observed scores are not as amenable to comparisons). Any case more than 3 standard deviations from the mean was considered an outlier, recoded as missing, and subsequently excluded from the analyses.

For high school GPA and standardized test scores, there were several steps taken to reconcile differences in the type of data (e.g., SAT vs. ACT) and the source of the data (self-report vs. institutional report). Several standardized admissions and placement test scores were considered as assessments of cognitive ability: SAT total score, ACT composite score, Accuplacer scores, and COMPASS scores. In order to summarize these data into one score, the four tests were concorded onto a common scale using previous research (see ACT, 2010; The College Board, 2009; Ellis, n.d.). As Dorans (2004a) pointed out, concordance is not intended to treat test scores as interchangeable measures of the same construct. Instead, these concorded scores "can be thought of as occupying the same location in a rank ordering of scores in some group of people" (p. 229). In other words, these scores are concorded in order to gather a general indicator of students' cognitive ability as it relates the population of college-going students.

For both test scores and high school GPA, institutional reports were used when available, but for ACT and SAT, self-reports were used if no other data were present. When multiple tests were available, ACT and SAT scores were used over COMPASS or Accuplacer, given some concern about students' perceptions of and efforts toward placement tests (Rosenbaum, Deil-Amen, & Person, 2006; Rosenbaum, Schuetz, & Foran, 2010). Similarly, preference was given to institutional reports of high school GPA. When institutional reports were not available, self-reported high school GPA values were included. For the remainder of this report, these aggregated variables will be referred to as *test scores* and *HSGPA*, respectively.

Analysis

The general procedure consisted of two steps. In the first step, the focus was on the incremental validity of SuccessNavigator over test scores and HSPGA. The second step consisted of determining the relative weights of test scores, HSGPA, and the subskills to form each index. Overall bivariate correlations between all 12 possible predictors—test scores, HSGPA, and the 10 SuccessNavigator subskill scores—and each of the four outcomes are provided in Table 7.

Table 7

Bivariate Correlation Between Predictors and Outcomes

	College GPA	Persistence	College-level math grade	College-level English grade
n	4,091	3,932	1,049	1,090
Test score	.359**	.106**	.169**	.275**
HSGPA	.377**	.156**	.325**	.336**
Meeting class expectations	155**	.053**	.081**	.176*
Organization	.146**	.035*	.124**	.133*
Commitment to college	.104**	.060**	.050	.117*
Institutional commitment	.000	.005	-.007	.053*
Stress sensitivity	.008	-.027	-.047	.000
Test anxiety	.022	-.037*	-.058	-.024
Academic self-efficacy	.143**	.049**	.050	.127**
Connectedness	.029	.002	.022	.128**
Institutional support	.022	-.011	.023	.071**
Barriers to success	.141**	.087**	.078*	.132**

Note. GPA = grade point average. HSGPA = high school grade point average.

** $p < .01$. *$p < .05$.

Incremental validity. The first step consisted of showing the extent to which SuccessNavigator scores contribute to the prediction of the outcome measures beyond what HSGPA and test scores predict. This was done by conducting a hierarchical regression analysis and comparing the R^2 values of three regression models. The first model included only test scores as a predictor; in the second model HSGPA was added; and finally, the subskills were added in the third model. If the model that included the subskills showed a significant change in R^2, it was concluded that the SuccessNavigator added valuable information in the prediction of the outcome variables.

Weighting predictor variables. After determining if the SuccessNavigator scores showed incremental validity in the prediction of the outcome variables, final weights were determined from an ordinary least squares multiple regression analysis. Since retention was measured as a binary variable, logistic regression was used in the formation of the retention success index.

Once again, a two-sample method was used to determine model strength and generalizability, using the same split that was used in the previous analyses. The first sample was used to determine the weights of the SuccessNavigator subskills and the R^2 values. The second sample was used to determine if the initial findings were comparable across samples.

When used in practice, the success indices are designed to optimally weight test scores, HSGPA, and SuccessNavigator scores. However, it is often the case that a given student could be missing test scores or HSGPA. For example, many institutions, such as community colleges, do not require standardized test scores for admissions. Additionally, students who were home-schooled or who received a GED would not have high-school grade information. Thus, we also explored scenarios when HSGPA or test scores were missing so that the indices could be applied when one of these variables was unavailable or not applicable.

Results

In examining the bivariate correlations in Table 7, we first see that test scores and HSGPA both have predictive efficacy across all four outcomes. Several psychosocial measures also have statistically significant relationships across outcomes, although some (e.g., academic self-efficacy) are significant for a subset of the outcomes. The low magnitude of some effect sizes (i.e., bivariate correlations) may be due to several factors. For one, these are first semester outcomes, which may be less reliable indicators of student success than those measured over a full year or more. Additionally, some of these factors (e.g., test anxiety) may have nonlinear effects that are not well-represented by traditional correlations. Overall, these findings support the use of SuccessNavigator for predicting academic outcomes, yet a regression-based approach, which allows for the examination of unique effects and relative contribution (i.e., comparing the contributions of psychosocial factors to those of test scores and HSGPA) provides further information. It was this methodology that was used in developing the success indices.

Academic success index. A total of 4,091 students had available first semester college GPA data and were thus included in the analysis. In this stage, data were only available for eight institutions, since one school could not provide outcomes data at the time of analysis. Table 8 shows the distribution of students by institution, gender, and race/ethnicity. Only sample sizes for the largest three ethnic groups are presented.

Table 8

Sample Sizes by Institution, Gender, and Race/Ethnicity Used for Academic Success Index Analysis

Institution	*n*	Gender		Race/ethnicity		
		Male	Female	White	African American	Hispanic/ Latino
1 [a]	1,565	715	850	175	178	948
2 [b]	978	347	631	308	18	471
3 [a]	448	242	206	222	48	106
4 [b]	422	124	298	328	17	45
5 [a]	376	168	208	292	30	32
8 [b]	110	42	68	35	5	32
9 [b]	135	60	75	90	4	32
7 [b]	57	18	39	36	4	6
Total	4,091	1,716	2,375	1,486	304	1,672

[a] 2-year institution. [b] 4-year institution.

Table 9 shows the change in R^2 (ΔR^2) in the prediction of college GPA when SuccessNavigator scores are included in the analysis. It can be seen that the percentage of variance explained by the SuccessNavigator scores over and above test scores and HSGPA is statistically significant ($\Delta R^2 = .028$, $p < .05$), with SuccessNavigator accounting for an additional 3% of the variance. Comparatively, HSGPA increases the accounted for variance by 5% when controlling for test scores. Robbins et al. (2004) found that psychosocial skills accounted for an additional 4% of the variance in first year GPA, when controlling for socioeconomic status, test scores, and HSGPA. Therefore, we consider this increase in the accounted for variance is both statistically and practically significant.

Table 9

Incremental Validity of SuccessNavigator in the Prediction of Grade Point Average (GPA)

Model	R	R^2	ΔR^2
1. Test scores	.377	.142	.142***
2. Test scores + HSGPA	.437	.191	.049***
3. Test scores + HSGPA + SN	.468	.219	.028***

Note. HSGPA = high school GPA, SN = SuccessNavigator.

*** p value < .001.

The academic success index was developed by first creating a composite score using the final regression model (Number 3) presented in Table 9. Table 10 shows the generalization of the academic success index across samples. Note that there is relatively small shrinkage of the percentage of variance explained from the test sample (R^2 = .235) to the validation sample (R^2 = .201). In addition, Table 10 also shows two models that remove HSGPA and test scores, respectively. Again, these models were designed to consider scenarios when such data were not available. Although there is some loss in predictive validity, these models still account for a sizable amount of variance in college GPA.

Table 10

Proportion of Variance Explained in the Different Scenarios for the Academic Success Index

R^2	Complete data	Missing HSGPA	Missing test scores
Sample 1	.235	.188	.189
Sample 2	.201	.158	.169

Note. HSGPA = high school grade point average.

The regression weights for the final model are presented in Table 11. These regression weights are unstandardized, although all variables have been rescaled (M = 100, SD = 15) so that all regression weights are comparable. These scores were used to create a composite score (equivalent to the predicted college GPA), and then these composite scores were standardized to a scale with a mean of 100 and standard deviation of 15 (in the same fashion as the general and subskill scores). It is this rescaled composite that represents the final academic success index.

Table 11

Final Regression Weights for the Academic Success Index

	Complete data	Missing HSGPA	Missing test scores
Intercept	-2.055**	-1.4052**	-1.1262**
Test scores	.021**	.0299**	--
HSGPA	.020*	--	.0287**
Meeting class expectations	.005*	.0071**	.0059*
Organization	.007**	.0096**	.0054**
Commitment to college	.003	.0034	.0038
Institutional commitment	-.005*	-.0055**	-.0068**
Stress sensitivity	-.004*	-.0055**	-.0034
Test anxiety	-.003	-.0023	-.0006
Academic self-efficacy	.005*	.0050	.0081**
Connectedness	.002	-.0001	-.0017
Institutional support	-.005*	-.0070**	-.0088**
Barriers to success	.005**	.0060**	.0068**
R^2 sample 1	.235	.188	.189
R^2 sample 2	.201	.158	.169

Note. HSGPA = high school grade point average.

* *p* value < .05. ** *p* value < .01.

A normative approach was used to create groupings (i.e., cut scores) for the academic success index. Using only students with complete data in the SuccessNavigator subskills, HSGPA, and test scores, the sample was divided into three score bands: the lowest quartile, the the middle two quartiles, and the highest quartile. Table 12 shows the cut scores that define each band, as well as the actual mean college GPA within that group.

Table 12

Cutoff Scores for Creating Score Bands for the Academic Success Index

	Low	Medium	High
Index score criterion for band	< 90.00	>= 90.00 <= 109.66	> 109.66
Predicted GPA criterion for band	< 2.23	>= 2.23 <= 2.97	> 2.97
Mean GPA within band	1.90	2.59	3.33

Note. HSGPA = high school grade point average.

Figure 1 shows the frequency of the actual grades within each band, demonstrating that the three score bands discriminate performance in observed college GPA. That is, students in the high band have higher predicted college GPA scores than students in the middle and in the low bands. This was considered as evidence of the validity of the academic success index.

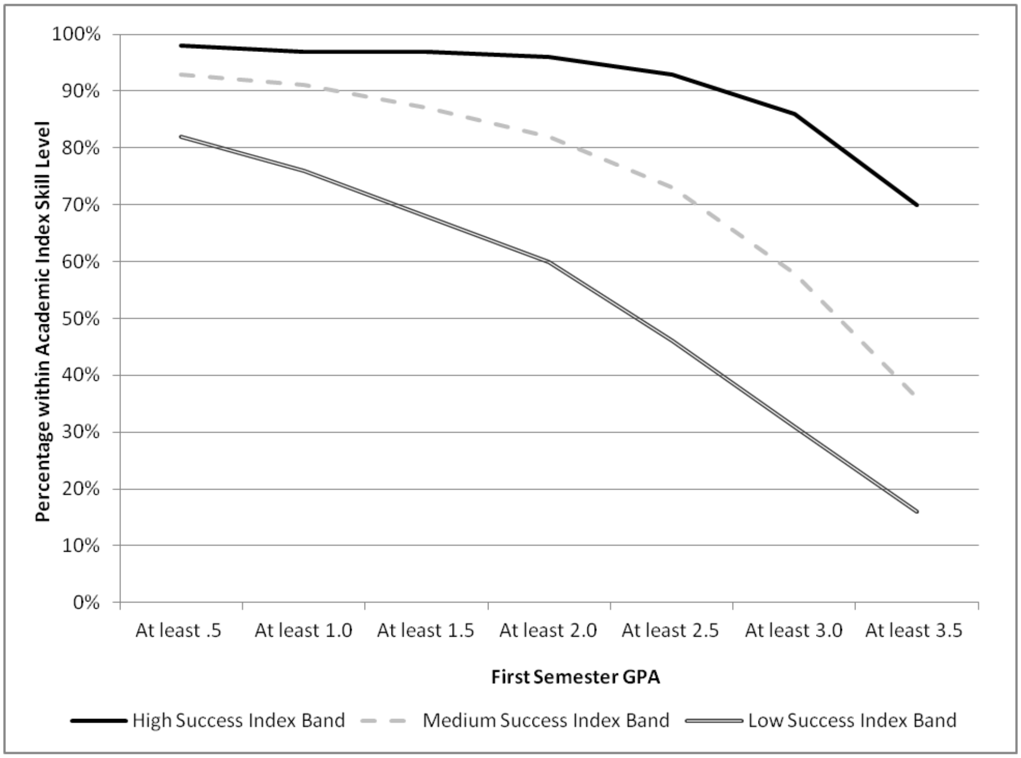

Figure 1. **Distribution of students by actual grade point average (GPA) and by academic success index score band.**

Retention success index. A total of 3,932 students had available persistence data and were thus included in the analysis. Table 13 shows the distribution of students by institution, gender, and race/ethnicity. Note that two additional institutions were excluded here. At the time of the analyses, persistence data were not available for students from Institution 7. Also, ancillary analyses within institutions showed an irregular pattern of results (Nagelkerke R^2 = 1.0) for Institution 8, which suggested issues due to the relatively small sample taken from this institution. Thus, these data were excluded from the analyses.

Table 13

Sample Sizes by Institution, Gender, and Race/Ethnicity Used for Retention Index Analysis

Institution	*n*	Gender		Race/ethnicity		
		Male	Female	White	African American	Hispanic/ Latino
1	1,566	716	850	175	178	948
2	985	349	636	310	19	473
3	448	242	206	222	48	106
4	422	124	298	328	17	45
5	376	168	208	292	30	32
9	135	60	75	90	4	32
Total	3,932	1,659	2,273	1,417	296	1,636

Retention was measured as a dichotomous variable that indicated whether or not students had registered for the following semester. A logistic regression was used to predict persistence from the 10 subskills, HSGPA, and test scores. The incremental validity results show that the contribution of SuccessNavigator scores is statistically significant over and beyond the effects of test scores and HSGPA, with effect sizes demonstrated by the change in Nagelkerke R^2 (Table 14). One note of caution is that the effect sizes in logistic regression, such as Nagelkerke R^2, cannot be interpreted as proportion of variance explained (see Nagelkerke, 1991), thus cross-validation and effect sizes for missing data models (similar to those presented in Table 10) are not presented.

Table 14

Incremental Validity of SuccessNavigator in the Prediction of Persistence

Model	Nagelkerke	ΔR^2
1. Test scores	.022	.022**
2. Test scores + HSGPA	.047	.025***
3. Test scores + HSGPA + SN	.074	.027***

Note. HSGPA = high school grade point average, SN = SuccessNavigator.

** *p* value < .01. *** *p* value < .001.

The creation of the retention success index was performed in the same manner as the academic success index. The regression weights presented in Table 15 were used to form composite scores, which were rescaled to have a mean of 100 and a standard deviation of 15. Once again, a normative approach was used to create the score bands for the retention success index. Using only students with complete SuccessNavigator, HSGPA, and test score data, the low, medium, and high score bands were formed by taking the bottom, middle two, and top quartiles, respectively. Table 16 shows the scores at the upper and lower quartiles, which define the three bands, as well as the actual persistence rates within each band. Figure 2 shows the persistence rates across the distribution of the retention index (separated into quartiles). The persistence rates across the three actual score bands, progressing from low to high, were 77.4%, 89.0%, and 94.2%.

Table 15

Final Logistic Regression Weights for the Academic Success Index

	Complete data	Missing HSGPA	Missing test scores
Intercept	-1.810**	-.7316	-.8511
Test scores	.0151**	.0250**	--
HSGPA	.0263**	--	.0303**
Meeting class expectations	.0123	.0140	.0129
Organization	-.0064	-.0029	-.0075
Commitment to college	.0082	.0095	.0084
Institutional commitment	-.0065	-.0088	-.0078
Stress sensitivity	-.0117*	-.0136*	-.0113
Test anxiety	-.0061	-.0049	-.0049
Academic self-efficacy	-.0058	-.0060	-.0034
Connectedness	.0027	.0056	.0028
Institutional support	-.0125	-.0143	-.0147**
Barriers to success	.0240*	.0247**	.0250**

Note. HSGPA = high school grade point average.

* *p* value < .05. ** *p* value < .01.

Table 16

Cutoff Scores for Creating Score Bands for the Retention Index

	Low	Medium	High
Index score criterion for band	< 90.26	>= 90.26 <= 109.72	> 109.72
Predicted persistence rate criterion for band	< 84.08	>= 84.08 <= 92.56	> 92.56
Mean persistence rate within band	77.42	88.96	94.20

Course acceleration. In developing and evaluating the course acceleration indices, a procedure described in Scott-Clayton (2012) was followed. This process determines the relationship between a set of predictors and success using only students registered in credit-bearing college-level courses. Then these regression models are fit onto students who placed into lower levels of coursework to estimate their likely success if placed into those college-level courses. Once again, indices were developed separately for mathematics and English using grades in those college-level courses as the dependent variable. Tables 17 and 18 show the distribution of the samples by institution, gender, and race used to create each index.[1]

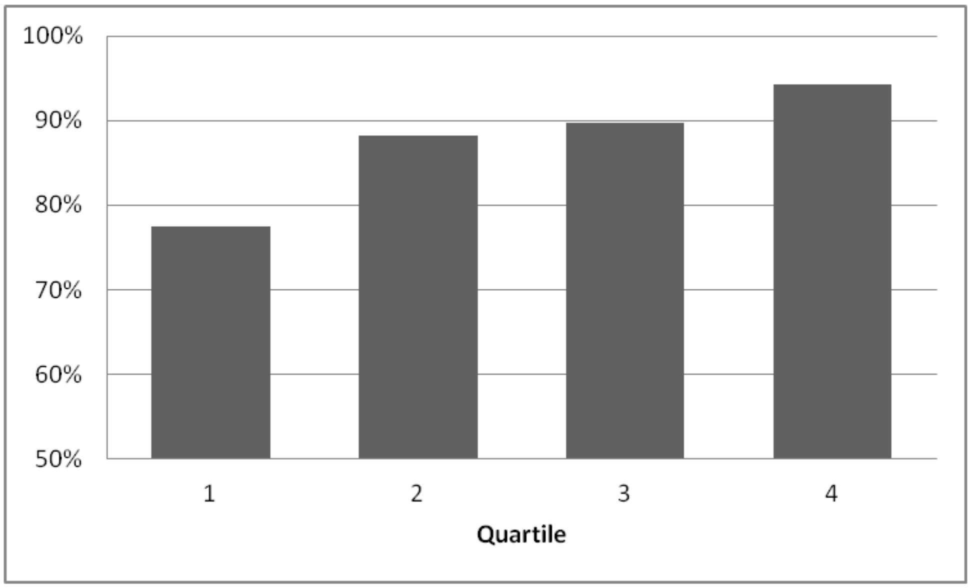

Figure 2. **Persistence rates by SuccessNavigator retention success index quartiles.**

Table 17

Sample Sizes by Institution, Gender, and Race/Ethnicity Used for Course Placement Index for Mathematics

		Gender		Race/ethnicity		
Institution	*n*	Male	Female	White	African American	Hispanic/Latino
2	641	245	396	224	14	273
4	169	68	101	142	4	10
9	89	44	45	58	1	24
3	74	42	32	46	6	10
5	64	43	21	54	2	5
1	53	21	32	10	5	27
Total	1,090	463	627	534	32	349

Table 18

Sample Sizes by Institution, Gender, and Race/Ethnicity Used for Course Placement Index for English

		Gender		Race/ethnicity		
Institution	*n*	Male	Female	White	African American	Hispanic/Latino
2	609	216	393	183	13	308
3	244	122	122	129	20	54
1	242	116	126	23	16	180
4	137	31	106	110	2	15
9	110	46	64	72	3	27
5	67	36	31	57	1	4
Total	1,409	567	842	574	55	588

Table 19 shows the R^2 values from the hierarchical models for English and mathematics scores. As stated, test scores were not included in the final course acceleration success indices, though they are presented in Table 19 for the purposes of demonstrating incremental validity. For each subject, models were built sequentially, including test scores in the first step, HSGPA in the second step, and SuccessNavigator subskill scores in the third step. The contribution of each model in the prediction of English and mathematics grades is indicated by the change in R^2 ($\triangle R^2$) between subsequent models. In both mathematics (R^2 = .023) and English (R^2 = .075), test scores alone significantly predicted course grades, though counting for a larger portion of variance in English.

Table 19

Incremental Validity of SuccessNavigator in the Prediction of Course Grades

Model	English			Mathematics		
	R	R^2	$\triangle R^2$	R	R^2	$\triangle R^2$
1. Test scores	.275	.075	.075***	.152	.023	.023***
2. Test scores + HSGPA	.369	.136	.061***	.307	.094	.071***
3. Test scores + HSGPA + SN	.407	.166	.029***	.337	.114	.020*

Note. HSGPA = high school grade point average, SN = SuccessNavigator.

* *p* value < .05. *** *p* value < .001.

Interestingly, these findings are counter to those of Scott-Clayton (2012), who found that placement tests were more predictive of grades in college-level mathematics (R^2 = .129) than English (R^2 = .017). These opposing findings may be due to the nature of our respective samples. Whereas our sample contains both 2- and 4-year institutions from varying settings across the United States, Scott-Clayton's sample consisted of only one large, urban community college.

Regardless, both sets of findings demonstrate differences in the phenomena underlying success in mathematics and English courses. Future research should explore these differences and the phenomena that underlie success in each subject.

HSGPA added significantly in both models, accounting for 6% and 7% of additional variance in English and mathematics, respectively. Finally, SuccessNavigator subskill scores accounted for significant variance in mathematics ($\triangle R^2$ = .020) and English ($\triangle R^2$ = .029), even when controlling for test scores and HSGPA. In total, the full models accounted for 11% of the variance in mathematics grades and 17% of the variance in English grades.

The full model, including test scores, HSGPA, and SuccessNavigator, is presented to demonstrate the incremental validity of SuccessNavigator. As previously discussed, the models used to develop the course acceleration indices, presented in Table 20, excluded tests scores. The weights from this modified model produced a composite score equivalent to predicted course grade in mathematics (predicted grade M = 2.25, SD = 0.52) and English (predicted grade M = 2.77, SD = 0.64; note that grade scales range from 0-4, with F = 0 and A = 4). As was the case with academic success and retention, these composites were then rescaled to have a mean of 100 and a standard deviation of 15 to form the course acceleration indices.

Table 20

Final Weights Used for Course Placement Indices for Mathematics and English

	English		Mathematics	
	Complete data	Missing HSGPA	Complete data	Missing HSGPA
Intercept	-1.170	.977*	-2.220**	1.929*
HSGPA	.031**	--	.045**	--
Meeting class expectations	.007	.012*	.002	.008
Organization	.000	.002	.002	.005
Commitment to college	.001	.004	.005	.004
Institutional commitment	-.005	-.008	.000	-.004
Stress sensitivity	-.003	-.006	-.007	-.010
Test anxiety	-.004	-.001	-.008	-.008
Academic self-efficacy	.004	.008	-.006	-.001
Connectedness	.011**	.014**	-.003	.000
Institutional support	-.006	-.010*	.003	-.002
Barriers to success	.004	.005	.011*	.013*
R^2 sample 1	.144	.056	.139	.043
R^2 sample 2	.129	.049	.076	.029

Note. HSGPA = high school grade point average.

* p value < .05. ** p value < .01.

In order to demonstrate additional validity of these models, we estimated over- and underplacement, mirroring the methodology presented in Scott-Clayton (2012) by comparing predicted and actual grades. *Underplacement* (a false-negative decision) refers to students who were placed into developmental courses, but were predicted to succeed (i.e., obtain a B or better) in the college-level course. Conversely, *overplacement* (a false-positive decision) refers to students who were placed into the college-level course, but were predicted to fail and did so.

The regression models represented in Table 19 (i.e., using test scores, HSGPA, and SuccessNavigator) were used to estimate predicted grades for students in college-level and developmental courses. For institutions with multiple levels of developmental courses, only students in the highest level (i.e., immediately below the college-level course) were included in the analysis. It was of special interest to determine the underplacement frequencies, given that SuccessNavigator focuses on accelerating students into higher level courses and is not designed to place students into lower levels of remediation.

Table 21 shows the frequencies of the actual and the predicted English grades of students in college and developmental courses. Note that the number of students who are predicted to have scores of A, B, and C, (n = 1,392 in the case with complete data; n = 1,409 in the case with missing HSGPA) is higher than the actual number of students with those grades (n = 1,181).

Table 21

Actual Versus Predicted Grades in English Courses

Actual grade		Predicted grades									
		Complete data					Missing HSGPA				
		F, I , W	D	C	B	A	F, I , W	D	C	B	A
College	A	0	3	43	451	93	0	0	45	528	17
	B	0	3	79	333	22	0	0	44	389	4
	C	0	6	49	95	4	0	0	32	120	2
	D	0	2	15	29	0	0	0	7	39	0
	F, I, W	1	2	66	109	4	0	0	42	140	0
Develop-mental	Pass	0	8	191	**396** [a]	**21** [a]	0	1	93	**516** [a]	**6** [a]
	Fail	0	2	48	99	2	0	0	22	126	3

Note. Bold = underplaced, I = incomplete, SGPA = high school grade point average, W = withdrawn.

[a] Indicates underplaced.

Regarding underplacement, Table 21 shows that 417 students with complete data and 522 students with missing HSGPA had a passing grade in English developmental courses and also were predicted to succeed (grade of A or B) in college level courses. Table 22 compares the actual and predicted grades for mathematics in college and developmental courses. For mathematics, 114 students with complete data and 260 with missing data in HSGPA were predicted to succeed in college level courses but were placed in developmental courses.

It is interesting to note that many students in Tables 21 and 22 were predicted to receive a C or higher, but actually received an F, incomplete, or withdrew from the course. One hypothesis is that many of the students who did not complete or withdrew from the course may have been able to successfully pass, but were otherwise unable. Future research should explore these false positives to improve the accuracy of predicting course grades.

Table 22

Actual Versus Predicted Grades in Mathematics Courses

Actual grade		Predicted grades									
		Complete data					Missing HSGPA				
		F, I , W	D	C	B	A	F, I , W	D	C	B	A
College	A	0	5	93	177	10	0	0	104	181	0
	B	1	12	138	151	3	0	0	144	160	1
	C	1	13	129	68	0	0	0	104	106	1
	D	0	6	47	20	1	0	0	38	36	0
	F, I, W	1	30	130	50	4	0	0	124	91	0
Develop-mental	Pass	2	49	263	**111** [a]	**3** [a]	0	1	167	**260** [a]	**0** [a]
	Fail	0	32	114	30	2	0	0	84	93	1

Note. I = incomplete, SGPA = high school grade point average, W = withdrawn.

[a] Indicates underplaced.

In creating scoring groups for the course acceleration success indices, a normative approach, similar to that used with the academic and retention success indices, was used. Again the sample was limited to students with complete data for the SuccessNavigator subskills and HSGPA. Here, only two score bands were created, rather than the three bands used for the other indices. The lower of the two bands consisted of students for whom acceleration should be considered with caution, since their predicted grades are in the lowest quartile. The higher band corresponded to the students in which course acceleration is recommended since their scores are above the lower quartile. Table 23 shows scale score and predicted grade at the cut point for each index, as well as the actual mean grade within each score band for both English and mathematics.

Table 23

Cutoff Scores for Creating Score Bands for the Course Placement Indices

	English		Mathematics	
	Caution	Accelerate	Caution	Accelerate
Index score range	< 91.38	>= 91.38	< 91.27	>= 91.27
Predicted grade range for band	< 2.48	>= 2.48	< 1.88	>= 1.88
Mean grade within band	2.23	2.82	1.26	2.40

Figures 3 and 4 demonstrate the distribution of the actual grades by placement band for English and mathematics. The low score band has a higher percentage of students with C, D, and F grades and a lower percentage of students with A grades. In contrast, the high score band has a lower percentage of students with C, D, and F students and a higher percentage of students with A grades.

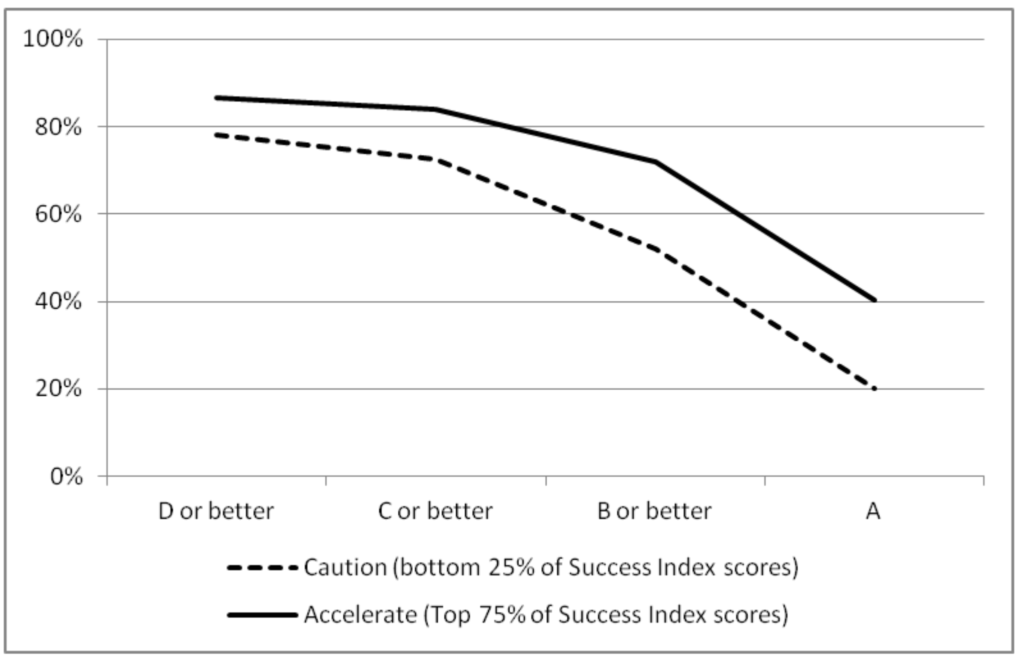

Figure 3. **Distribution of students in English courses by actual grade and by placement band.**

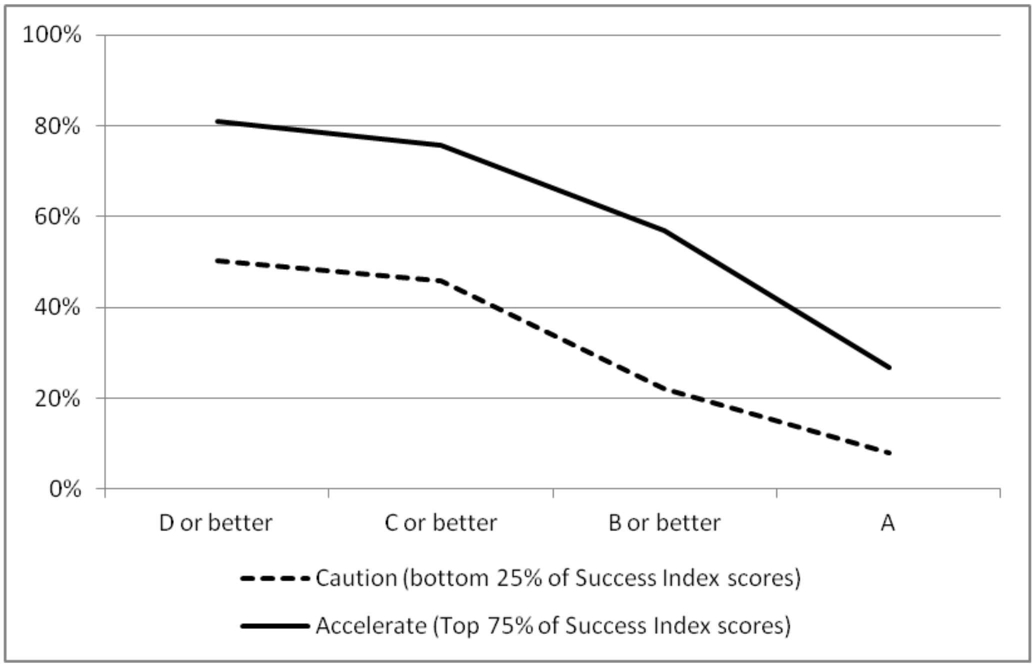

***Figure 4.* Distribution of students in mathematics courses by actual grade and by placement band.**

Fairness

Test fairness has become an increasingly important topic, and demonstrations of fairness are emphasized in both the *ETS Standards for Quality and Fairness* (Educational Testing Service [ETS], 2002) as well as the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). The ETS standards state that all materials and services must take the diversity of the populations served into account as well as remain mindful to include only construct-relevant information that appreciates the value of both the assessment and interpretation. Specifically, fairness emphasizes that:

> ….products and services will be designed, developed, and administered in ways that treat people equally and fairly regardless of differences in personal characteristics such as race, ethnicity, gender, or disability that are not relevant to the intended use of the product or service. (ETS, 2002, p. 17)

Thus, in addition to determining if an assessment's items and scales function properly among themselves, it is also important to ensure that the assessment functions equally well across important subgroups—namely gender and race/ethnicity. If it does not, then a test may possess bias that poses a threat to validity. The success indices presented by SuccessNavigator add an additional consideration of fairness, in that the inferences drawn about a student's likely success must be comparable across these subgroups.

In the cases of both measurement and prediction, invariance studies have been recommended to assess a test's fairness (Dorans, 2004b, Meredith, 1993). Invariance studies empirically test whether parameters differ significantly across two groups. When testing whether the structure determined by factor analyses is similar across two groups, the comparability of parameters is referred to as *measurement invariance*. When testing whether an assessment predicts a given outcome equally across two groups, the comparability of the parameters is referred to as *predictive invariance*. Accordingly, we conducted several procedures to examine if SuccessNavigator measures constructs similarly across gender and racial/ethnic subgroups and if the success indices predict relevant outcomes in the same way across these groups.

Measurement invariance. Subgroup means by gender and race/ethnicity for each subskill are shown in Table 24. In order to ensure that the subgroup mean differences found were due to differences in the population and not artifacts of the instrument, measurement invariance studies by gender and race/ethnicity were conducted in each subskill. Measurement invariance is established through several models, each of which differs in the extent to which the factor structure can vary across two groups. This progressive assessment of measurement invariance is conducted by comparing a series of nested models (Jöreskog, 1971; Sörbom, 1974; Vandenberg & Lance, 2000).

First, configural invariance—a basic test of measurement equivalence—simply tests the overall structure across groups, determining if the same items relate to the same factors in each group. This was examined by fitting a unidimensional model in all groups. Next, metric invariance is assessed by constraining the item loadings to have the same values across groups. Finally, strong factorial invariance was tested by additionally constraining the item intercepts to be the same across groups. The nested models were evaluated with the chi-square difference test. Since the chi-square difference test is highly sensitive to sample size, the CFI was also examined, with changes of 0.01 considered acceptable (Chen, West, & Sousa, 2006).

Table 24

Subskill Subgroup Means by Gender and Race

Subskill	Gender		Race		
	Male	Female	White	African American	Hispanic/ Latino
Meeting class expectations	4.79 (.62)	4.92 (.56)	4.95 (.55)	4.84 (.65)	4.81 (.59)
Organization	3.62 (.80)	4.12 (.79)	3.98 (.86)	3.89 (.78)	3.85 (.82)
Commitment to college	5.27 (.63)	5.39 (.50)	5.34 (.54)	5.30 (.62)	5.38 (.52)
Institutional commitment	4.99 (.74)	5.11 (.68)	5.06 (.71)	5.09 (.69)	5.12 (.65)
Sensitivity to stress	4.32 (.85)	3.97 (.86)	4.10 (.88)	4.21 (.90)	4.13 (.84)
Test anxiety	3.52 (1.07)	3.12 (1.00)	3.35 (1.03)	3.37 (1.06)	3.21 (1.05)
Academic self-efficacy	5.15 (.66)	5.23 (.59)	5.22 (.61)	5.14 (.72)	5.21 (.59)
Connectedness	4.26 (.87)	4.26 (.85)	4.32 (.86)	4.33 (.88)	4.23 (.83)
Institutional support	4.52 (.73)	4.61 (.70)	4.58 (.72)	4.70 (.69)	4.56 (.70)
Barriers to success	4.58 (.70)	4.68 (.66)	4.76 (.64)	4.58 (.81)	4.56 (.65)

Note. The standard deviations are shown in parenthesis below the subgroup mean.

In the cases in which the models of metric or strong factorial invariance did not fit the data, partial invariance models were examined. Partial invariance models are more realistic models in which some of the item parameters are constrained for invariance while others are allowed to vary between groups. Modification indices (MI) were examined to determine which item parameters needed to be freely estimated across groups. The content of the items that were found to have violations of metric or strong factorial invariance were examined to determine if there was bias in the items.

Sample. The structure of the subskills was examined using data in the sample used for selecting the items that consisted of a total of 5,061 college students for whom demographic information was available (Table 1). The same two sample approach that was used in the structural validity analyses was applied here.

Analysis. Multigroup analysis with respect to gender and race/ethnicity were conducted using M*plus* 6.1 (Muthén & Muthén, 1998–2010) to guarantee that the relationship between the latent variable and the items was similar across groups.

***Results: Measurement invariance by gender*.** The data from 2,078 female and 2,983 male responses were used. The fit of the models testing for configural, metric, and strong factorial invariance for each subskill across gender are presented in Table 25. This table also indicates the items that were freely estimated to achieve partial invariance, where applicable. Ultimately, partial strong factorial invariance was achieved with respect to gender in all subskills.

***Results: Measurement invariance by race/ethnicity*.** Data from 1,911 White, 408 African American, and 1,923 Hispanic students were analyzed for invariance across racial/ethnic subgroups. It should be noted that while the sample sizes for White and Hispanic students are comparable, the sample size for African American students is roughly one fourth of that size, thus results should be interpreted with caution. Table 26 shows the model fit statistics for each model by race/ethnicity and subskill, including the items that were freely estimated in an ethnic group, where applicable. The results indicate that, in all scales, it was possible to achieve partial factorial invariance.

Interpreting measurement invariance results. Tables 25 and 26 show that partial factorial invariance with respect to gender and race/ethnicity was achieved in every subskill. Thus, while some items were invariant, other items were freely estimated across groups. This is a common finding in psychological testing, where factorial invariance in all items is not frequently found (Schmitt & Kuljanin, 2008). The question that remains deals with the practical significance of partial invariance, that is, the impact of partial invariance on the conclusions reached with the test. Unfortunately, there are no clear guidelines to evaluate when the violations of measurement invariance can be considered negligible and when they become large. For example, while Reise, Widaman, and Pugh (1993) have argued that the majority of the items have to be invariant to ensure meaningful group comparisons, Steenkamp and Baumgartner (1998) have suggested that only one invariant item is needed in addition to the referent indicator. Another complication is that violations of invariance at the item level do not necessarily mean violations to invariance at the scale level (Stark, Chernyshenko, & Drasgow, 2004). The measurement invariance results for SuccessNavigator suggest that group mean differences should be taken with caution in the subskills with large number of noninvariant items as indicated in Tables 25 and 26 (e.g., organization).

Table 25

Model Fit for Measurement Invariance Models by Gender in Each Subskill

Subskill	Model	RMSEA	CFI	χ^2	df	$\Delta\chi^2$	Δdf	$\Delta\chi^2$ p value	Items parameters freely estimated
Meeting class expectations	Configural	.045	.982	351.78	58	--	--	--	
	Metric	.042	.982	365.47	66	13.27	8	.103	BC12
	Strong	.042	.982	369.81	69	4.76	3	.190	BC12, BC10, BC7, BC1, BC5, BC2
Organization	Configural	.029	.995	94.16	30	--	--	--	
	Metric	.028	.995	98.69	33	4.53	3	.210	Org3, Org2, Org6, Org8, Org11
	Strong	.027	.995	100.67	35	1.97	2	.370	Org3, Org2, Org6, Org8, Org11, Org10r
Commitment to college	Configural	.072	.961	762.01	54	--	--	--	
	Metric	.068	.960	770.89	61	8.88	7	.261	DG4
	Strong	.064	.961	779.08	68	8.19	7	.316	DG4
Institutional commitment	Configural	.055	.987	333.94	38	--	--	--	
	Metric	.052	.987	341.46	44	7.52	6	.276	Inc7
	Strong	.049	.987	350.04	49	8.58	5	.127	Inc7, Inc2
Sensitivity to stress	Configural	.038	.995	112.61	28	--	--	--	
	Metric	.031	.996	124.87	36	12.27	8	.134	Str12r,
	Strong	.030	.996	135.45	42	10.57	6	.102	Str12r, Str1r, Str2r

Subskill	Model	RMSEA	CFI	χ^2	df	$\Delta\chi^2$	Δdf	$\Delta\chi^2$ p value	Items parameters freely estimated
Test anxiety	Configural	.057	.975	497.25	54	--	--	--	
	Metric	.053	.975	509.44	62	12.19	8	.143	
	Strong	.052	.975	510.63	65	11.84	3	.757	Tdur10r, Tdur6r, Tdur5r,Tdur 4, Tpst5r
Academic self-efficacy	Configural	.039	.991	201.20	42	--	--	--	
	Metric	.037	.990	208.41	47	7.21	5	.206	ASE6r, ASE5, ASE8
	Strong	.036	.990	212.73	50	4.33	3	.228	ASE6r, ASE5, ASE8, ASE2, ASE9
Connectedness	Configural	.051	.987	211.30	28	--	--	--	
	Metric	.047	.987	214.39	33	3.10	5	.685	CNCT2
	Strong	.045	.987	218.09	36	3.70	3	.296	CNCT2, CNCT4, CNCT10
Institutional support	Configural	.044	.982	430.15	72	--	--	--	
	Metric	.043	.982	443.78	79	13.63	7	.058	NSUP12r, NSUP6r, NSUP13r
	Strong	.042	.983	448.91	83	5.13	4	.274	Nsup12r, Nsup6r, Nsup13r, NSup5r, Nsup4, Nsup3
Barriers to success	Configural	.044	.984	268.78	46	--	--	--	
	Metric	.041	.984	281.92	54	13.13	8	.107	Strm5, Strm6
	Strong	.038	.984	288.05	61	6.13	7	.529	Strm5, Strm6, Strm10r

Note. CFI = comparative fit index, RMSEA = root mean square error approximation.

Table 26

Model Fit for Measurement Invariance Models by Race/Ethnicity in Each Subskill

Subskill	Model	RMSEA	CFI	χ^2	df	$\Delta\chi^2$	Δdf	$\Delta\chi^2$ p value	Freely estimated items parameters		
									White	African American	Hispanic/ Latino
Meeting class expectations	Configural	.041	.985	294.25	87	--	--	--			
	Metric	.038	.985	321.77	105	27.53	18	.070			
	Strong	.037	.984	338.57	117	16.79	12	.158	BC4, BC12, BC10	BC7, BC8, BC3	
Organization	Configural	.036	.993	126.98	45	--	--	--			
	Metric	.032	.993	143.99	59	17.01	14	.256	Org10r	Org5r	
	Strong	.030	.992	158.36	70	14.37	11	.213	Org10, Org7r, Org11,	Org10	Org10
Commitment to college	Configural	.080	.953	811.46	81	--	--	--			
	Metric	.075	.952	833.68	94	22.22	13	.052	DG10, DG3. DG12		
	Strong	.071	.952	850.68	104	17.00	10	.074	DG10, DG3. DG12, DG7, DG11	DG7	DG7
Institutional commitment	Configural	.053	.988	286.10	57	--	--	--			
	Metric	.048	.987	303.82	71	17.72	14	.220			
	Strong	.045	.987	319.59	82	15.77	11	.150	Inc2, Inc9, Inc7		
Sensitivity to stress	Configural	.040	.995	137.50	42	--	--	--			
	Metric	.035	.995	157.34	58	19.84	16	.227	STRS15, STRS3		
	Strong	.033	.994	180.66	72	23.33	14	.055	STRS15, STRS3, STRS1r		STRS9r

Subskill	Model	RMSEA	CFI	χ^2	df	$\Delta\chi^2$	Δdf	$\Delta\chi^2$ p value	Freely estimated items parameters		
									White	African American	Hispanic/ Latino
Test anxiety	Configural	.055	.978	422.55	81	--	--	--			
	Metric	.050	.978	440.53	96	17.99	15	.263	Tpre5r		
	Strong	.048	.978	456.03	106	15.49	10	.115	Tpre5r, Tdur4r, Tpst5r tdur7r, Tdur5r	Tdur8r	
Academic self-efficacy	Configural	.040	.990	208.05	63	--	--	--			
	Metric	.037	.989	228.77	78	20.73	15	.146		ASE2	
	Strong	.037	.989	241.32	90	12.55	12	.403	ASE12r, ASE6r	ASE2	ASE9
Connectedness	Configural	.057	.984	238.29	42	--	--	--			
	Metric	.052	.983	258.49	54	20.20	12	.063			
	Strong	.049	.983	268.95	62	10.46	8	.234	CNCT1, CNCT5, CNCT7		CNCT10
Institutional support	Configural	.049	.978	481.88	108	--	--	--			
	Metric	.045	.978	499.95	128	18.07	20	.582			
	Strong	.043	.978	518.10	141	18.15	13	.152	NSUP13r	NSUP3	NSUP5r, NSUP1, NSUP3, NSUP2, NSUP10
Barriers to success	Configural	.046	.983	273.56	69	--	--	--			
	Metric	.041	.983	302.22	89	28.67	20	.095			
	Strong	.039	.982	318.12	101	15.90	12	.196	Strm2r, Strm1r, Strm10r	Strm6, Strm3r	Strm9r, Strm7, Strm4r

Note. CFI = comparative fit index, RMSEA = root mean square error approximation.

Differential prediction. One approach to judge the implications of violations of invariance is to consider the purpose of the test. It has been argued that achieving factorial measurement invariance is fundamental when the test is used to compare group means and for selection (Borsboom, 2006). SuccessNavigator, however, was not designed to compare groups in specific subskills. Rather, prediction via the success indices is a key intended use. Accordingly, in order to determine fairness, it is most relevant to examine the differential impact of the SuccessNavigator scores in the prediction of these various outcomes across gender and race/ethnicity. This shifts the focus from measurement invariance to predictive invariance, or differential prediction.

While measurement invariance examined whether there is systematic error in the measure of a group, differential prediction is not concerned with the evaluation of the test itself, but rather the function of the test in predicting a criterion (Meade & Fetzer, 2009). Differential prediction of the SuccessNavigator subskills was examined using the same four different criteria that are used in the success indices: overall college GPA, retention, and grades in English and mathematics in a college-level course.

Analysis**.** Once again, all analyses were conducted in M*plus* 6.1 (Muthén & Muthén, 1998–2010). The method for evaluating differential prediction consisted of conducting a regression analysis in which the 10 SuccessNavigator subskill scores, HSGPA, and test scores were used as predictors of the outcome variables. First, a model in which the slopes were constrained to be the same across subgroups was evaluated using the χ^2 of model fit. A significant χ^2 indicated that a model with equal slopes across groups did not fit the data. A nonsignificant χ^2 was taken as evidence that slope invariance holds in the data.

If slope invariance was found, the intercepts were constrained to be the same across groups. A χ^2 difference test was again used to compare the fit of a model with constraints in the slopes and intercepts to a model with constraints in the slopes only. A significant χ^2 would indicate that the fit of the model significantly decreased by constraining the intercepts to equality; hence, the intercepts were considered to be different across groups. In contrast, a nonsignificant χ^2 difference test indicated that the intercepts were the same across groups. Invariance in the slopes and intercepts is referred to as regression invariance.

Results: College GPA. A total of 1,502 males and 2,111 females were included in the differential prediction analysis by gender; responses from 1,322 White, 260 African American, and 1,490 Hispanic students were included in the differential prediction analysis by race/ethnicity.

The fit of the regression models are shown in Table 27. Males and females were found to have the same regression slopes and intercepts in the prediction of college GPA, indicating regression invariance. The results for race/ethnicity indicate that there is slope invariance but that there are group differences in the intercepts. Meade and Fetzer (2009) delineated the reasons why there could be differences in the intercepts. One of the reasons they mentioned and that might explain the results found here is differences in the criterion. Mean college GPA for White students is 3.03, while the mean for African American students is 2.49, and the mean for Hispanic students is 2.73. The regression intercepts for these groups were -.705, -.881 and -.772, respectively. Thus, the differences in the college GPA of the three ethnic groups are most likely responsible for the differences in intercepts. That is, the relationship between the SuccessNavigator indices and success is similar across groups, though the groups have different levels of predicted success that likely are attributable to differences in actual success.

Table 27

Fit of the Models Evaluating Predictive Invariance for Grade Point Average (GPA)

Group	Model	χ^2	df	*p* value	RMSEA	CFI	$\Delta\chi^2$	Δdf	$\Delta\chi^2$ *p* value
Gender	Equal slopes	17.883	12	0.1193	0.016	0.993	--	--	--
	Equal slopes and intercepts	18.913	13	0.1258	0.016	0.993	1.03	1	0.310
Race	Equal slopes	16.373	24	0.8742	0	1	--	--	--
	Equal slopes and intercepts	26.325	26	0.4454	0.003	0.999	9.952	2	0.007

Note. CFI = comparative fit index, RMSEA = root mean square error approximation.

Results: Retention. A total of 1,659 males and 2,273 females were analyzed for differential prediction of retention. For differential prediction by race/ethnicity, data from 1,417 White students, 296 African American students, and 1,636 Hispanic students were analyzed. Because the outcome variable was dichotomous, the χ^2 difference test was conducted using the Satorra-Bentler scaled χ^2. This was done in M*plus* by using the DIFFTEST option. Results indicate that regression invariance was found with respect to gender and race (see Table 28).

However, this does not mean that the groups have similar rates of persistence; it simply means that students with the same set of characteristics from different gender or racial/ethnic groups will have similar accuracy of prediction.

Table 28

Fit of the Models Evaluating Predictive Invariance for Retention

Group	Model	χ^2	df	p value	$\Delta\chi^2$	Δdf	$\Delta\chi^2$ p value
Gender	Equal slopes	8.425	12	0.7511	--	--	--
	Equal slopes and intercepts	9.775	13	0.7122	0.066	1	0.797
Race	Equal slopes	14.409	24	0.9369	--	--	--
	Equal slopes and intercepts	19.30	26	0.8535	1.183	2	0.553

***Results: Grades in English*.** Data from 567 males and 842 females enrolled in college level English courses were analyzed for gender predictive invariance. Because data from only 55 African Americans were available, race/ethnicity comparison were conducted only with data from 574 White students and 588 Hispanic students.

The results indicate that slope invariance was achieved for gender, but that a model with equal intercepts did not fit the data. As in the case for college GPA, group differences in the criterion are the likely explanation of the differences in intercepts. The mean of English grades for females was 3.0, while for males, 2.6 (regression intercepts of -2.066 and -2.310, respectively). Regression invariance was, however, achieved for race/ethnicity (see Table 29). Once again, though, this does not indicate similarity in group mean grades; it simply indicates equivalence of prediction.

Table 29

Fit of the Models Evaluating Predictive Invariance for English Grades

Group	Model	χ^2	df	p value	RMSEA	CFI	$\Delta\chi^2$	Δdf	$\Delta\chi^2$ p value
Gender	Equal slopes	5.868	12	0.9226	0	1	--	--	--
	Equal slopes and intercepts	16.186	13	0.2393	0.019	0.985	10.32	1	0.001
Race	Equal slopes	17.369	12	0.1362	0.028	0.973	--	--	--
	Equal slopes and intercepts	18.934	13	0.1251	0.028	0.97	1.56	1	0.211

Note. CFI = comparative fit index, RMSEA = root mean square error approximation.

Results: Grades in mathematics. Data from 463 males and 627 females were analyzed for differential prediction regarding mathematic grades in college courses. Once again, African Americans were excluded from the race/ethnicity analyses due to limited sample size (n = 32). Thus, differential prediction by race/ethnicity was conducted only with data from 534 White students and 349 Hispanic students. Table 30 shows that regression invariance was achieved for both gender and race/ethnicity, indicating no evidence of differential prediction of mathematics grades across these subgroups. Again, this does not speak to group mean similarity; it merely indicates similarity in predictive validity across groups.

Table 30

Fit of the Models Evaluating Predictive Invariance for Mathematics Grades

Group	Model	χ^2	df	p value	RMSEA	CFI	$\Delta\chi^2$	Δdf	$\Delta\chi^2$ p value
Gender	Equal slopes	14.647	12	0.2613	0.02	0.997	--	--	--
	Equal slopes and intercepts	15.678	13	0.267	0.019	0.976	1.031	1	0.310
Race	Equal slopes	12.612	12	0.3979	0.011	0.993	--	--	--
	Equal slopes and intercepts	13.063	13	0.443	0.964	0.999	0.451	1	0.502

Note. CFI = comparative fit index, RMSEA = root mean square error approximation.

Conclusions about predictive invariance. No evidence of differential prediction was found for retention and for mathematic grades with respect to gender and race. In the prediction of college GPA, regression invariance was found with respect to gender and slope invariance was found with respect to race. In the prediction of English grades, regression invariance was found for race and slope invariance was found with respect to gender. Although significant differences in the intercepts were found in the prediction of mathematics GPA by race and in the prediction of English grades by gender, the differences in intercepts were small in practical terms. Further, because slope invariance was found with respect to the four outcome variables, it can be concluded that the SuccessNavigator subskills have similar impacts across subgroups.

Discussion

This report has presented a great deal of information about the reliability of SuccessNavigator scores, the validity of the inferences that can be drawn from SuccessNavigator, and the equality of measurement and prediction for SuccessNavigator scores across gender, racial, and ethnic subgroups. Here, we briefly summarize and clarify these findings.

Overall, SuccessNavigator scores are highly reliable. The extent to which scores are replicable or stable is a critical feature of any assessment, as scores that contain high amounts of random error inherently lack validity. Our results found that all but one subskill scale exceeded .8, with the lone exception achieving $\alpha = .78$, surpassing accepted benchmarks for reliability in a low-stakes testing context. The general skill scores showed even higher levels of reliability.

SuccessNavigator is strongly supported by theory and practice, providing evidence of content or substantive validity. Here, we reviewed the way that personality theory informed the development of the SuccessNavigator general skills and subskills. In addition, we discussed its alignment to existing practices in higher education, as well as consultations with students, faculty, staff, and administrators that supported its design.

Factor analyses empirically support the relationships among SuccessNavigator items and scores of SuccessNavigator, providing evidence of structural validity. By conducting factor analyses of the SuccessNavigator subskills, we were able to identify the items that best represent each construct and determine that the measurement models adequately fit the data. Moreover, relationships among the subskill scores support the general skills and the overall structure of SuccessNavigator.

Results from the Success Indices show that SuccessNavigator significantly contributes to the prediction of college GPA, persistence, and course grades even when controlling for standardized test scores and high school GPA, which provides evidence of predictive (or external) validity. Using sequential regression models, we were able to demonstrate the predictive value of SuccessNavigator for a host of educational outcomes. This supports the use of the success indices for identifying students' likelihood of success and, when coupled with a traditional placement test, recommending them for acceleration within or out of developmental mathematics and English courses.

Invariance studies show that SuccessNavigator predicts equally well across gender, racial, and ethnic subgroups, supporting the fairness of the assessment. Given the intended use of the assessment, predictive invariance studies showed that SuccessNavigator predicts college GPA, persistence, and course grades equally well across all relevant subgroups.

Future Directions

Indeed, supporting the use of a test with reliability and validity evidence is not a single event, but an ongoing process. Although this report supports the potential value of SuccessNavigator, there are still many questions to be explored with this assessment. The key validity concern for any placement decision is whether students who are accelerated are ultimately more successful in their college coursework than students of comparable abilities (both cognitive and noncognitive) who are first placed into a lower level remedial course. Students with relatively low probability of success, if placed directly into the more advanced class, may or may not be more likely to succeed in that course if they are first placed into a remedial course. This is an answerable empirical question, but not one that could be addressed in this study.

SuccessNavigator is a tool that is designed to serve the heterogeneous population of students in postsecondary education. Although we have addressed issues of measurement and prediction with some groups here, future research should continue to determine the performance of this assessment for different populations. We mentioned the relatively small sample of African American students in this study. Future research should explore this population of students, intentionally oversampling if necessary. Yet gender and race/ethnicity are only two of many important population characteristics. Given a rapidly changing and diversifying system of higher education, studies should also be conducted to examine the performance of SuccessNavigator with students who either demonstrate nontraditional characteristics (e.g., previous graduates who are returning for retraining) or are enrolled in nontraditional institutions, such as online or competency-based programs.

Continued study should also monitor the way SuccessNavigator predicts and impacts other educational outcomes. Although first semester college GPA and persistence were used here, the ultimate goal is to use SuccessNavigator to help students persist to completion. Accordingly, we should examine the way in which the use of SuccessNavigator relates to changes in rates of success. That is, when institutions adopt SuccessNavigator and use its scores to facilitate interactions with students, do rates of success—whether it is academic, retention, or classroom success—increase. This again raises the question of consequential validity (Messick, 1994, 1995; Moss, 1992). In other words, whether or not a test relates to a given outcome (predictive validity), can its implementation and use relate to change in that outcome?

Conclusion

As both students and institutions of higher education seek to improve success and degree attainment, psychosocial skills are a critical piece of the puzzle that has largely been unaddressed. Although increased attention is paid to the way students study, their classroom behavior, their goals and motivations, their ability to handle stress, and their ability to use resources, students and educators alike can benefit from a quality assessment that provides feedback on such factors.

This report has provided a sound basis of support for SuccessNavigator. Through examining several aspects of reliability, validity, and fairness, this appears to be an assessment that holds promise in helping students and those that work with them.

References

ACT. (2010, February). *Concordant ACT, COMPASS, and ASSET scores.* Iowa City, IA: Author. Retrieved from http://www.act.org/compass/pdf/Concordance.pdf

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment.* Washington, DC: U.S. Department of Education.

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college.* Washington, DC: U.S. Department of Education.

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House, 78*(5), 218-223.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17,* 10-17.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1-40.

Borsboom, D. (2006). When does measurement invariance matters? *Medical Care, 44,* 176-181.

Boylan, H. R. (2009). Targeted intervention for developmental education students (T.I.D.E.S.). *Journal of Developmental Education, 32*(3), 14-23.

Bridgeman, B. (1974). Effects of test score feedback on immediately subsequent test performance. *Journal of Educational Psychology*, *66*(1), 62.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123-143.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage Publications.

Burke, B. L. (2006). For the 'grader' good. *Observer, 19*(11), 33-37.

Camara, W. J., & Echternacht, G. (2000, July). *The SAT I and high school grades: Utility in predicting success in college* (Research Notes No. 10). New York, NY: The College Board.

Chemers, M. M., Hu, L. T., & Garcia, B. F. (2001). Academic self-efficacy and first year college student performance and adjustment. *Journal of Educational Psychology*, *93*(1), 55.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second order models of quality of life. *Multivatiate Behavioral Research*, *4*, 189-225.

The College Board. (2009, October). *ACT and SAT concordance tables* (Research Notes No. RN-40). New York, NY: Author. Retrieved from http://research.collegeboard.org/sites/default/files/publications/2012/7/researchnote-2009-40-act-sat-concordance-tables.pdf

Complete College America. (2012). *Remediation: Higher education's bridge to nowhere.* Washington, DC: Author.

Conley, D. T. (2007). *Toward a more comprehensive conception of college readiness.* Eugene, OR: Educational Policy Improvement Center.

Council for the Advancement of Standards in Higher Education. (2011). *The role of counseling services.* Washington, DC: Author. Retrieved from http://collegecounseling.org/docs/CAS.pdf

Dorans, N. J. (2004a). Equating, concordance, and expectation. *Applied Psychological Measurement*, *28*(4), 227-246.

Dorans, N. J. (2004b). Using subpopulation invariance to assess test score equity. *Journal of Educational Measurement*, *41*(1), 43-68.

Educational Testing Service. (2002). *ETS standards for quality and fairness.* Princeton, NJ: Author.

Ellis, S. F. (n.d.). *Correlation between various placement instruments for reading language/writing mathematics elementary algebra*. . Retrieved from the West Virginia Department of Education web site: http://wvde.state.wv.us/abe/documents/CorrelationBetweenVarious PlacementInstruments.pdf

Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes* (Research and Occasional Paper Series No. CSHE.6.07). Berkeley, CA: Center for Studies in Higher Education. Retrieved from http://cshe.berkeley.edu/publications/docs/ROPS.GEISER._SAT_6.12.07.pdf

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409-426.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a Common Core construct? *Journal of Personality and Social Psychology, 83,* 693-710.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education/Praeger.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*, 1-73.

Karabenick, S. A. (2003). Seeking help in large college classes: A person-centered approach. *Contemporary Educational Psychology*, *28*(1), 37-58.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006, July). *What matters to student success: A review of the literature. Commissioned report for the National Symposium on Postsecondary Student Success: Spearheading a dialog on student success*. Retrieved from http://nces.ed.gov/npec/pdf/kuh_team_report.pdf

Liu, O. L., Rijmen, F., MacCann, C., & Roberts, R. (2009). The assessment of time management in middle-school students. *Personality and Individual Differences, 47*(3), 174-179.

MacCann, C., Duckworth, A. L., & Roberts, R. D. (2009). Empirical identification of the major facets of conscientiousness. *Learning and Individual Differences*, *19*(4), 451-458.

Mattern, K. D., & Packman, S. (2009). *Predictive validity of ACCUPLACER scores for course placement: A meta-analysis* (College Board Research Report No. 2009-2). New York, NY: The College Board.

Mattern, K. D., & Patterson, B. F. (2009). *Is performance on the SAT related to college retention?* (College Board Research Report No. 2009-7). New York, NY: The College Board.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 51,* 81-90.

Meade, A. W., & Fetzer, M. (2009). Test bias, differential prediction, and a revised approach for determining the suitability of a predictor in a selection context. *Organizational Research Methods, 12*, 738-761.

Mehrens, W. A. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis, 6*(14). Retrieved from http://epaa.asu.edu/ojs/article/download/580/703

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58*(4), 525-543.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13-23.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*(4), 5-8.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62,* 229-258.

Muthén, L. K., & Muthén, B. O. (1998–2010). M*plus* (Version 6.1) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78,* 691-692.

Napoli, A. R., & Wortman, P. M. (1996). A meta-analysis of the impact of academic and social integration of persistence of community college students. *Journal of Applied Research in the Community College, 4*(1), 5-21.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

O’Banion, T. (2013). *Academic advising: The key to student success*. Washington, DC: American Association of Community Colleges.

Porchea, S. F., Allen, J., Robbins, S., & Phelps, R. P. (2010). Predictors of long-term enrollment and degree outcomes for community college students: Integrating academic, psychosocial, socio-demographic, and situational factors. *The Journal of Higher Education, 81*, 750-778.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135* (2), 322-338.

Reckase, M. D. (1998). Consequential validity from the test developer's perspective. *Educational Measurement: Issues and Practice*, *17*(2), 13-16.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552-566.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353-387.

Robbins, S. B., Allen, J., Casillas, A., Akamigbo, A., Saltonstall, M., Campbell, R., & Gore, P. A. (2009). Associations of resource and service utilization, risk level, and college outcomes. *Research in Higher Education, 50*, 101-118.

Robbins, S. Allen, J., Casillas, A., Peterson, C., & Le, H. (2006). Unraveling the differential effects of motivational and skills, social, and self-management measures from traditional predictors of college outcomes. *Journal of Educational Psychology, 98,* 598-616.

Robbins, S., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130* (2), 261-288.

Roberts, R. D., Schulze, R., & MacCann, C. (2007, April). *Student 360^{TM}: A valid medium for noncognitive assessment?* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Rosenbaum, J., Deil-Amen, R., & Person, A. (2006). *After admission: From college access to college success*. New York, NY: Russell Sage Foundation.

Rosenbaum, J., Schuetz, P., & Foran, A. (2010, July). *How students make college plans and ways schools and colleges could help* (Working paper). Evanston, IL: Northwestern University, Institute for Policy Research.

Sawyer, R. (2010). *Usefulness of high school average and ACT scores in making college admission decisions* (ACT Research Report Series No. 2010-2). Iowa City, IA: ACT, Inc.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*, 210-222.

Scott-Clayton, J. (2012). *Do high-stakes placement exams predict college success?* (CCRC Working Paper No. 41). New York, NY: Columbia College Research Center. Retrieved from http://ccrc.tc.columbia.edu/publications/high-stakes-placement-exams-predict.html?UID=1026

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27,* 229-239.

Sorey, K., & Duggan, M. (2008). Differential predictors of persistence between community college adult and traditional-aged students. *Community College Journal of Research and Practice, 32*(2), 75-100.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the effects of differential item (functioning and differential) test functioning on selection decisions: When are statistically significant effects practically important? *Journal of Applied Psychology, 89*, 497-508.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.

Trapmann, S., Hell, B., Hirn, J. W., & Schuler, H. (2007). Meta-analysis of the relationship between the Big Five and academic success at university. *Journal of Psychology, 215*, 132-151.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4-70.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 380-392). New York, NY: Guilford Press.

Wiley, A., Wyatt, J., & Camara, W. J. (2010). *The development of a multidimensional college readiness index* (College Board Research Report No. 2010-3). New York, NY: The College Board. Retrieved from http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2010-3-development-multidimensional-college-readiness-index.pdf

Notes

[1] Preliminary analyses revealed a zero correlation between HSGPA and mathematics grades for students at Institution 1. This was surprising given that HSGPA has been shown to be an important predictor of grades (see Geiser & Santelices, 2007). Further consultation with the institution revealed that HSGPA is actually used in placement decisions, thus creating a restriction of range that could bias the results for the general population (which does not typically use HSGPA). Thus, data from Institution 1 were excluded from the analysis.

Appendix

General skill	Subskill	Definition	Example item
Academic skills : Tools and strategies for academic success	Organization	Strategies for organizing work and time	I write a daily to-do list. I use a calendar to plan my school day.
	Meeting class expectations	Doing what’s expected to meet the requirements of courses including assignments and in-class behaviors.	I am on time for class. I complete my assignments on time.
Commitment: active pursuit toward an academic goal	Commitment to college goals	Perceived value and determination to succeed in and complete college	One of my life goals is to graduate college. The benefit of a college education outweighs the cost.
	Institutional commitment	Attachment to and positive evaluations of the school	This is the right school for me. I’m proud to say I attend this school.
	Sensitivity to stress	Tendency to feel frustrated, discouraged, or upset when under pressure or burdened by demands	I get stressed out easily when things don't go my way. I am easily frustrated.
	Academic self-efficacy	Belief in one’s ability to perform and achieve in an academic setting	I'm confident that I will succeed in my courses this semester. I can do well on tests if I apply myself.
	Test anxiety	General reactions to test-taking experiences, including negative thoughts and feelings (e.g., worry, dread)	When I take a test, I think about what happens if I don't do well. The night before a test, I feel troubled.
	Connectedness	A general sense of belonging and engagement	I feel connected to my peers. People understand me.
Social support: Connecting with people and student resources for success	Institutional support	Attitudes about and tendency to seek help from established resources	If I don't understand something in class, I ask the instructor for help. I know how to find out what's expected of me in classes.
	Barriers to success	Financial pressures, family responsibilities, conflicting work schedules, and limited institutional knowledge	Family pressures make it hard for me to commit to school. People support me going to college.