RESEARCH REPORT 2012-5

CollegeBoard

The Validity of the Academic Rigor Index (ARI) for Predicting FYGPA

By Krista D. Mattern and Jeffrey N. Wyatt

Krista D. Mattern is an associate research scientist at the College Board.

Jeffrey N. Wyatt is an associate research scientist at the College Board.

Acknowledgements

The authors wish to acknowledge Andrew Wiley and Paul Sackett for reviewing earlier versions of this manuscript. We would also like to thank Brian Patterson for his assistance with SAS programming and related questions.

Mission Statement

The College Board's mission is to connect students to college success and opportunity. We are a not-for-profit membership organization committed to excellence and equity in education.

About the College Board

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of more than 5,900 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT® and the Advanced Placement Program®. The organization also serves the education community through research and advocacy on behalf of students, educators and schools.

For further information, visit www.collegeboard.org.

VALIDITY

For more information on College Board research and data, visit www.collegeboard.org/research.

Contents

Tables

Executive Summary

A recurrent trend in higher education research has been to identify additional predictors of college success beyond the traditional measures of high school grade point average (HSGPA) and standardized test scores, given that a large percentage of unaccounted variance in college performance remains. A recent study by Wyatt, Wiley, Camara, and Proestler (2012) expanded the definition of college readiness beyond test scores and HSGPA to include a measure of the academic rigor or challenge associated with a student's course work in high school, referred to as the academic rigor index (ARI). This study represents the first examination of the validity of ARI in predicting first-year grade point average (FYGPA). The correlation between ARI and FYGPA indicated a moderate effect overall and by gender, ethnicity, and household income subgroups; however, ARI did not add incremental validity above SAT scores and HSGPA. Additionally, when added to SAT scores and HSGPA, ARI had no impact on differential prediction by relevant subgroups. Given the current movement toward a more holistic assessment of college applicants, a standardized measure of the academic rigor of a student's course load in high school suggests a promising additional measure to the assessment of a student's level of college readiness.

Introduction

Traditionally, when examining the relationship between the academic preparedness of high school seniors and their subsequent college performance, the focus has often been limited to the evaluation of SAT® scores and HSGPA (Bridgeman, McCamley-Jenkins, & Ervin, 2000; Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008; Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008; Patterson, Mattern, & Kobrin, 2009; Patterson & Mattern, 2011). Both test scores and HSGPA have large correlations with college grades; however, a considerable amount of variance in FYGPA remains unexplained. As such, a recurrent trend in higher education research has been to identify additional predictors of college success beyond the traditional measures of HSGPA and standardized scores (e.g., Camara & Kimmel, 2005; Willingham & Breland, 1982).

Academic Rigor

One such additional predictor has been the academic rigor or difficulty of high school course work. The importance of academic rigor during high school has been widely discussed at the national level. Specifically, in 1983, the National Commission of Excellence in Education released the seminal work *A Nation at Risk*, which indicated that U.S. students were not academically prepared for life after high school. The report found that approximately 13% of 17-year-olds were functionally illiterate and that the percentage for minority students could have been three times as high. Furthermore, the commission found, as the U.S. was slipping into educational mediocrity, other nations were forging ahead. The report's comparisons of the performance of U.S. students to that of students from other industrialized nations on 19 academic tests revealed that U.S. students never had the highest or even the second-highest performance, but rather had the lowest performance on seven tests.

The Commission identified "content" or curriculum as one source of the educational deficiencies of U.S. students. Specifically, "twenty-five percent of the credits earned by general track high school students are in physical and health education, work experience outside the school, remedial English and mathematics, and personal service and development courses, such as training for adulthood and marriage" (National Commission on Excellence in Education, 1983, p. 21). Based on these findings, the Commission recommended more rigorous high school graduation requirements, which they called the "New Basics." The highest level of the New Basics consisted of four years of English, three years of mathematics, three years of science, three years of social studies, and half a year of computer science. Additionally, for students who planned to attend an institution of higher education, the Commission recommended at least two years of a foreign language. One limitation of the New Basics was that it only addressed the *number* of courses in a given content area, whereas follow-up research also took into account the highest *level* of a course (Adelman, 1999; 2006; Horn & Kojaku, 2001).

Adelman (1999) created a measure of *academic curriculum intensity* that took Carnegie units as well as the content or level of the courses (e.g., highest mathematics course) into account. Carnegie units are based on time in the classroom whereby a yearlong course translates to 1.0 Carnegie units. There are 31 levels of the academic intensity variable with requirements for the highest level as follows: (1) 3.75 or more Carnegie units of English; (2) 3.75 or more Carnegie units of mathematics; (3) a highest mathematics course of either calculus, precalculus, or trigonometry; (4) 2.5 or more Carnegie units of science or more than 2.0 Carnegie units of core laboratory science (biology, chemistry, and physics); (5) more than 2.0 Carnegie units of foreign languages; (6) more than 2.0 Carnegie units of history and/or social studies; and (7) no remedial English or mathematics. In 2006, Adelman added a computer science course as a requirement.

Academic Rigor and College Readiness

Several studies have suggested that academic rigor is an important component of success in college. Adelman, Daniel, and Berkovits (2003) found a negative relationship between academic rigor in high school and remediation in college. Adelman (2006) found that academic intensity or academic rigor had the strongest relationship with bachelor's degree completion, more so than HSGPA or standardized test scores. The powerful relationship between rigorous course participation in high school and college success can be seen by examining the relationship between the most advanced mathematics course completed and college graduation. Approximately 83% of 12th-graders who had taken a calculus course in 1992 graduated with a bachelor's degree by 2000 compared to 75% for precalculus, 60% for trigonometry, and 40% for Algebra II. Such findings may explain the reason that many measures of college readiness include academic rigor or course participation as one component of the overall measure (Berkner & Chavez, 1997; Greene & Winters, 2005).

Several studies have suggested that academic rigor is an important component of success in college.

Academic Rigor Index (ARI)

One of the obstacles to creating and utilizing a comprehensive measure of academic rigor is that it is traditionally calculated from high school transcripts; this is a labor-intensive process that is very difficult to do on a large scale. To address this problem, Wiley, Wyatt, and Camara (2010) used student responses on the SAT Questionnaire — a survey students complete during registration for the SAT — about the courses they have completed during high school (see also Wyatt et al., 2012). To construct the ARI, courses were coded into five subscales: (1) English, (2) mathematics, (3) science, (4) social sciences and history, and (5) foreign and classical languages. For each subscale, a student can earn up to five points, with points awarded based on the association between course participation and FYGPA in college. Each subscale has its own set of parameters in terms of what earns a student more points; however, in general, students earn more points if they: (1) take more years or courses in that subject area; (2) take more rigorous courses in a subject area (e.g., honors, AP®, dual

… the focus of the current study is to fill this research gap, and measure the predictive validity or correlation between ARI and FYGPA using the same procedures used to calculate the predictive validity of HSGPA and SAT scores.

enrollment); or (3) take higher-level courses within a natural progression (for example, a student who has progressed through calculus will receive more points than a student who has progressed through Algebra II). The ARI scale is the sum of all five subscales; therefore, the score scale ranges from 0 to 25. For more specific details about the development and scoring of the ARI, refer to Wyatt et al. (2012).

Descriptive analyses have suggested that ARI scores were positively associated with college enrollment, FYGPA, and retention, thus supporting prior research findings that suggested that academic rigor is an important component of college success. However, validity evidence on the ARI has thus far been limited solely to such descriptive analyses. As such, the focus of the current study is to fill this research gap, and measure the predictive validity or correlation between ARI and FYGPA using the same procedures used to calculate the predictive validity of HSGPA and SAT scores.

Current Study

Published nearly three decades ago, *A Nation at Risk* (National Commission on Excellence in Education, 1983) represents the commencement of a national dialogue on improving the rigor of U.S. education that is ongoing today. Many resources are now devoted to the development of the Common Core State Standards, an initiative to determine what students need to know and be able to do to be successful in college and work (National Governors Association Center for Best Practices and the Council of Chief State School Officers, 2010). Given that this issue remains at the forefront of educational concerns, this study is timely. It will contribute to the literature by being the first, to our knowledge, to address four research gaps in the academic rigor literature. Employing the ARI as a standardized measure of academic rigor, this study examined (1) the validity of the academic rigor for predicting FYGPA; (2) whether academic rigor exhibits differential validity by gender, ethnicity, and household income; (3) whether academic rigor provides any increment in validity, above and beyond HSGPA and SAT scores, in the prediction of FYGPA; and (4) whether differential prediction is reduced when academic rigor is added to the model that already includes HSGPA and SAT scores.

Method

Sample

College performance data were obtained from a partnership between the College Board and 129 four-year institutions that agreed to provide college performance data (e.g., course grades, FYGPA, and retention) on their 2008 class of entering first-year students for research and validation purposes. These data were matched back to College Board records to obtain official SAT scores and responses to the SAT Questionnaire, which include self-reported HSGPA and demographic information. The sample of institutions was diverse with respect to region of the U.S., control (i.e., public versus private), selectivity, and size (refer to Patterson & Mattern, 2011, for more details about the sample). The final sample included 145,131 students.

Measures

ARI. The ARI is a composite measure of the level of difficulty or rigor associated with students' high school course work. The index is calculated from student responses to the SAT Questionnaire, which records information on English, mathematics, science, social science and history, and foreign and classical language courses completed during high school. In addition, students indicated the academic level of each course completed, such as honors, dual enrollment, or AP.

Within each subject area (English, mathematics, science, the social sciences and history, and foreign and classical languages), students are awarded between 0 and 5 points depending on the rigor of the student's course work. Each of the scores from these five subscales is summed, yielding a total score on a 0–25 scale. The algorithm for the scale and subscales was largely empirically based, and was derived by evaluating the relationship between course work and FYGPA. The complete algorithm is presented in Wiley et al. (2010).

SAT Scores. SAT scores were obtained for each sample of students for the critical reading (SAT-CR), mathematics (SAT-M), and writing (SAT-W) sections. Each section has a score scale range of 200 to 800 with 10-point increments.

HSGPA. Cumulative HSGPA was self-reported by students on the SAT Questionnaire (SAT-Q), which is completed during registration for the SAT. Grades were reported in letter grades ranging from an F (below 65) to an A+ (97–100).

Household Income. Household income was obtained from self-reported data on the SAT-Q .

Gender. Students provided gender information (female or male) when they completed the SAT-Q.

Ethnicity. Students indicated their race/ethnicity on the SAT-Q. The categories include (1) Native American or Alaska Native; (2) Asian, Asian American, or Pacific Islander; (3) black or African American; (4) Mexican or Mexican American; (5) Puerto Rican; (6) other Hispanic, Latino, or Latin American; (7) white; and (8) other. In this report, categories 4, 5, and 6 were combined into a single category titled "Hispanic."

FYGPA. FYGPA was obtained from participating colleges and universities. The values of FYGPA ranged from 0.00 to 4.07, with only one student having a FYGPA greater than 4.00.

Data Analysis

Descriptive Statistics. Means and standard deviations for SAT, HSGPA, and ARI were computed for the total sample and by gender, race/ethnicity, and household income.

Predictive Validity. To assess the validity of the ARI for predicting FYGPA, correlations were computed between ARI, SAT scores, and HSGPA with FYGPA within each institution. All correlations were corrected for restriction of range in the predictor variables at the institution level using the Pearson–Lawley multivariate correction, with the 2008 College-Bound Seniors cohort serving as the population (Gulliksen, 1950; Lawley, 1943). At this point, the corrected covariance matrices for each institution were averaged, weighting by the number of students included for analysis of that institution. To estimate the multiple correlations, the corrected sample size–weighted average correlation matrix was used to compute multiple correlations; this procedure was described in Powers (2004).

To examine the extent to which ARI exhibits differential validity across student subgroups for FYGPA, correlations were computed for each subgroup within each institution. Institutions that had fewer than 15 students of a particular subgroup were excluded from that group's results. For example, if an institution had fewer than 15 males (e.g., a single-sex institution), then that institution was excluded from the male analyses. However, if the same institution had at least 15 females, it would be included in the analyses for females. These raw covariance matrices were separately corrected at the level of the institution using the

Pearson–Lawley multivariate correction for restriction of range, with the 2008 College-Bound Seniors cohort serving as the population (Gulliksen, 1950; Lawley, 1943). The covariance matrices were then transformed into correlation matrices; after they were weighted by subgroup-by-institution sample size, the average of the corrected correlation matrices was treated as the pooled, subgroup-specific correlation matrix. On the basis of this matrix, multiple correlations were computed.

Finally, the issue of whether the ARI provided incremental validity above and beyond traditional admission criteria of SAT scores and HSGPA in the prediction of FYGPA was examined. Specifically, the multiple correlation of SAT scores and HSGPA with FYGPA was compared to the multiple correlation of SAT scores, HSGPA, and ARI with FYGPA. If the multiple correlation of SAT scores, HSGPA, and ARI with FYGPA is greater than the multiple correlation of SAT scores and HSGPA with FYGPA, it would indicate that ARI provides incremental validity to the prediction of FYGPA.

Differential Prediction. Differential prediction occurs when a test systematically over- or underpredicts the criterion (e.g., FYGPA) by student subgroups. This is calculated by subtracting the predicted FYGPA derived from a regression analysis from the earned FYGPA (i.e., residual = $FYGPA_{earned} - FYGPA_{predicted}$). Negative values (residuals) indicate overprediction, and positive values indicate underprediction. For example, if a specific subgroup (e.g., females) tends to earn higher FYGPAs than is predicted by a regression equation using ARI, then the ARI exhibits differential prediction by gender, namely underprediction for females.

To assess the extent to which the ARI, as well as SAT and HSGPA, exhibits differential prediction, regression equations within the institution were estimated. The average residual of FYGPA by various student subgroups was computed across the entire sample for various predictor sets. Specifically, regression analyses were run with each predictor included in the model separately, as well as a model with SAT and HSGPA and a model with all three measures (ARI, SAT, and HSGPA).

Results

Descriptive Statistics

Performance on each of the study variables was calculated by gender, race/ethnicity, and household income (refer to Table 1). The results by gender reveal that males had higher SAT-CR and SAT-M scores, whereas females had higher SAT-W, HSGPA, ARI scores, and FYGPA. As for the racial/ethnic subgroups, white students had the highest SAT-CR and FYGPA, and Asian American students had the highest SAT-M, SAT-W, HSGPA, and ARI. African American students had the lowest mean performance on all five academic measures. As for the results by household income, a positive relationship between each of the academic measures and income is apparent. Specifically, higher income categories were associated with higher means.

Predictive Validity

The main goal of this study was to examine the predictive validity of ARI in terms of FYGPA. In the original study, Wiley et al. (2010) provided descriptive statistics in terms of mean FYGPA by ARI scale point as well as the ARI score associated with a 65% probability of earning

Table 1.

Means and Standard Deviations of Study Variables by Demographic Characteristics

		SAT-CR		SAT-M		SAT-W		HSGPA		ARI		FYGPA	
Subgroup	***N***	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Gender													
Female	80,666	550	93	554	92	552	93	3.66	0.47	13.4	5.3	3.06	0.68
Male	64,465	561	95	596	94	544	94	3.57	0.50	13.2	5.3	2.90	0.74
Ethnicity													
American Indian	724	545	90	550	91	526	84	3.56	0.49	12.2	5.1	2.84	0.70
Asian American	14,484	555	100	616	96	559	100	3.68	0.44	14.6	5.4	3.03	0.67
African American	9,990	494	90	494	89	487	88	3.41	0.55	11.2	5.2	2.59	0.78
Hispanic	12,785	518	91	530	90	510	89	3.58	0.49	13.0	5.3	2.76	0.76
White	100,385	565	90	579	90	558	90	3.63	0.48	13.4	5.3	3.05	0.69
Other	3,659	556	94	569	96	553	96	3.59	0.48	13.4	5.3	2.98	0.69
No Response	3,104	583	98	586	96	573	98	3.64	0.49	13.7	5.4	3.04	0.71
Household Income													
< $20,000	5,459	493	97	515	101	488	93	3.54	0.53	11.5	5.3	2.71	0.80
$20,000–$40,000	11,340	518	93	536	96	509	90	3.59	0.52	12.3	5.3	2.79	0.79
$40,000–$60,000	14,097	538	92	552	93	528	91	3.61	0.50	12.7	5.2	2.89	0.75
$60,000–$80,000	15,666	547	91	562	93	537	90	3.62	0.50	12.9	5.3	2.96	0.73
$80,000–$100,000	15,390	558	90	574	92	549	89	3.63	0.48	13.3	5.3	3.00	0.70
$100,000–$120,000	13,289	563	89	582	89	556	88	3.64	0.48	13.6	5.2	3.04	0.69
$120,000–$140,000	6,816	566	88	584	88	559	88	3.63	0.48	13.8	5.2	3.04	0.68
$140,000–$160,000	5,077	571	88	589	88	566	87	3.63	0.47	13.9	5.2	3.05	0.67
$160,000–$200,000	5,977	572	88	591	89	568	89	3.62	0.48	13.8	5.3	3.07	0.66
> $200,000	9,441	582	86	603	86	583	88	3.59	0.47	14.1	5.3	3.10	0.62
No Response	42,579	566	94	584	96	562	95	3.63	0.47	13.7	5.3	3.05	0.69
Total	145,131	555	94	573	95	549	94	3.62	0.49	13.3	5.3	2.99	0.71

a B- (2.67) in the first year of college; however, no correlational analyses were conducted. Therefore, this study represents the first examination of the predictive validity, differential validity, and incremental validity of the ARI. Results are provided in Table 2.

Overall, the corrected correlation between ARI and FYGPA was .44 (observed = .25), which is considered a medium effect. SAT and HSGPA were more strongly correlated with FYGPA than ARI, with corrected correlations of .55 and .56, respectively. The multiple correlation of SAT and HSGPA with FYGPA was .63. The change in the multiple correlation due to the inclusion of the ARI was .00, indicating that ARI does not provide any incremental validity over and above SAT scores and HSGPA in the prediction of FYGPA. The reason that ARI may not account for any additional variance in FYGPA is likely a function of its high correlation with both SAT scores (r = .66) and HSGPA (r = .50). Refer to the appendix for more information on the corrected and observed correlations among the predictors.

Correlations were also computed by gender, race/ethnicity, and household income. As was the case for SAT and HSGPA, the correlation between ARI and FYGPA was slightly higher for females (.45) compared to males (.43). The correlations between each of the three academic measures and FYGPA were highest for Asian American and white students. As for household income, the correlation between each of the academic measures and FYGPA tended to be higher for higher-income categories; however, the differences were small. Finally, similar to the overall results, ARI failed to provide any incremental validity over and above SAT scores and HSGPA for any of the subgroup analyses, with one exception. For American Indian students, the inclusion of ARI increased the multiple correlation from .51 to .52, or a change of .01, although the sample size for the group was relatively small and should be interpreted with caution.

Table 2.

Corrected (Observed) Correlations of ARI, SAT, and HSGPA with FYGPA by Demographic Characteristics

		N	ARI	SAT	HSGPA	SAT & HSGPA	SAT, HSGPA, ARI
Gender	Female	80,666	.45 (.24)	.58 (.41)	.56 (.37)	.65 (.48)	.65 (.48)
	Male	64,465	.43 (.25)	.53 (.35)	.54 (.37)	.61 (.45)	.61 (.46)
Ethnicity	American Indian	335	.42 (.25)	.45 (.33)	.44 (.29)	.51 (.40)	.52 (.40)
	Asian American	14,174	.43 (.19)	.54 (.33)	.54 (.30)	.61 (.42)	.61 (.42)
	African American	9,721	.39 (.21)	.47 (.28)	.46 (.31)	.53 (.38)	.53 (.39)
	Hispanic	12,564	.38 (.18)	.49 (.30)	.49 (.31)	.56 (.39)	.56 (.39)
	White	100,368	.44 (.25)	.54 (.35)	.58 (.40)	.64 (.47)	.64 (.47)
	Other	3,182	.39 (.20)	.51 (.36)	.49 (.30)	.58 (.43)	.58 (.43)
	No Response	2,690	.40 (.22)	.52 (.37)	.51 (.35)	.59 (.45)	.59 (.45)
Household Income	< $20,000	5,056	.37 (.19)	.46 (.32)	.47 (.31)	.53 (.40)	.53 (.40)
	$20,000–$40,000	11,193	.40 (.22)	.49 (.34)	.51 (.36)	.58 (.44)	.58 (.44)
	$40,000–$60,000	14,002	.41 (.23)	.51 (.36)	.54 (.39)	.61 (.46)	.61 (.46)
	$60,000–$80,000	15,612	.43 (.25)	.52 (.36)	.56 (.40)	.62 (.47)	.62 (.47)
	$80,000–$100,000	15,291	.44 (.25)	.54 (.37)	.56 (.39)	.63 (.47)	.63 (.47)
	$100,000–$120,000	13,143	.43 (.25)	.53 (.36)	.58 (.40)	.64 (.47)	.64 (.47)
	$120,000–$140,000	6,513	.42 (.23)	.54 (.36)	.57 (.39)	.63 (.47)	.63 (.47)
	$140,000–$160,000	4,703	.45 (.27)	.55 (.37)	.58 (.40)	.65 (.49)	.65 (.49)
	$160,000–$200,000	5,575	.42 (.24)	.52 (.34)	.56 (.37)	.62 (.45)	.62 (.45)
	> $200,000	9,229	.42 (.23)	.52 (.32)	.57 (.38)	.63 (.45)	.63 (.45)
	No Response	42,579	.44 (.25)	.56 (.39)	.57 (.38)	.65 (.48)	.65 (.48)
Total		145,131	.44 (.25)	.55 (.37)	.56 (.38)	.63 (.47)	.63 (.47)

Note: Correlations are computed within the institution. Correlations were corrected for range restriction using the Pearson–Lawley multivariate correction (Gulliksen, 1950; Lawley, 1943; Pearson, 1902). Average sample size weighted correlations are shown in parentheses. SAT is the multiple correlation for all three sections.

Differential Prediction

The final set of analyses examined whether ARI systematically over- or underpredicted FYGPA for gender, racial/ethnic, and household income subgroups, and if so, how this over- or underprediction compared to the differential prediction findings of the SAT and HSGPA. Additionally, the extent to which the inclusion of ARI reduced the magnitude of over- and underprediction compared to the model that used only SAT and HSGPA was examined. As summarized in Table 3, HSGPA resulted in the least amount of differential prediction by gender, followed by ARI and SAT scores. Though the magnitude of the prediction error varied slightly for the three measures, the direction of error was the same across all three. Specifically, FYGPA was underpredicted for females and overpredicted for males regardless of which measure was employed. When used in combination, the magnitude of differential prediction by gender was the same whether or not ARI was included. That being said, based on a standard deviation of around .70 for FYGPA, the magnitude of prediction error was small (i.e., all *ds* ≤ |.15|; Cohen, 1988) regardless of the model evaluated.

The differential prediction analyses by ethnicity reveal that ARI results in the most differential prediction of the three academic measures, followed closely by HSGPA. SAT resulted in the least amount of differential prediction, especially for African American and Hispanic students, compared to that of ARI and HSGPA. Specifically, all three measures overpredicted FYGPA for African American and Hispanic students, but the magnitude of overprediction was larger for ARI and HSGPA. When all three measures were included in the model, the magnitude of differential prediction by ethnicity was essentially equivalent to the model that used only SAT scores and HSGPA. The one exception was for African American students, for whom overprediction of FYGPA increased slightly, from -.11 to -.12 with the inclusion of ARI. In other words, adding ARI to HSGPA and SAT scores did not result in more or less differential prediction by ethnicity.

As for the differential prediction results by household income, all three measures overpredicted FYGPA for students from low-income families ($60,000 or less); however, the SAT scores resulted in the least amount of differential prediction, followed by ARI and HSGPA. Adding ARI to the model that included SAT scores and HSGPA had no impact on the differential prediction by income categories; in fact, the last two columns are identical.

Table 3.

Average Overprediction (-) and Underprediction (+) of FYGPA by ARI, SAT, and HSGPA

		N	ARI	SAT	HSGPA	SAT & HSGPA	SAT, HSGPA, ARI
Gender	Female	80,666	0.07	0.08	0.05	0.06	0.06
	Male	64,465	-0.09	-0.11	-0.06	-0.07	-0.07
Ethnicity	American Indian	724	-0.12	-0.09	-0.12	-0.09	-0.09
	Asian American	14,484	-0.01	0.00	0.01	0.00	0.00
	African American	9,990	-0.26	-0.14	-0.23	-0.11	-0.12
	Hispanic	12,785	-0.15	-0.06	-0.14	-0.06	-0.06
	White	100,385	0.05	0.02	0.04	0.02	0.02
	Other	3,659	-0.01	0.00	0.01	0.01	0.01
	No Response	3,104	0.01	-0.03	0.02	-0.02	-0.02
Household Income	< $20,000	5,459	-0.15	-0.05	-0.18	-0.07	-0.07
	$20,000–$40,000	11,340	-0.12	-0.05	-0.14	-0.07	-0.07
	$40,000–$60,000	14,097	-0.05	-0.02	-0.07	-0.04	-0.04
	$60,000–$80,000	15,666	0.00	0.01	-0.02	-0.01	-0.01
	$80,000–$100,000	15,390	0.02	0.01	0.01	0.00	0.00
	$100,000–$120,000	13,289	0.04	0.02	0.04	0.02	0.02
	$120,000–$140,000	6,816	0.03	0.01	0.04	0.02	0.02
	$140,000–$160,000	5,077	0.03	0.00	0.04	0.01	0.01
	$160,000–$200,000	5,977	0.04	0.01	0.06	0.04	0.04
	> $200,000	9,441	0.02	-0.02	0.07	0.03	0.03
	No Response	42,579	0.03	0.01	0.04	0.02	0.02
Total		145,131	0.00	0.00	0.00	0.00	0.00

Note: Regression analyses were run within the institution, and mean residuals were computed over the total sample by subgroup.

Discussion

The correlational analyses reveal that ARI not only is indicative of future performance as signified by its positive relationship with FYGPA, but also allows students to showcase their strengths and abilities via a new dimension: the rigor of high school courses completed. This measure has also been shown to have smaller subgroup differences than two of the most traditionally employed cognitive measures of academic preparation: standardized test scores and high school grades (Wiley et al., 2010). Given the current movement toward a more holistic assessment of college applicants, a standardized measure of the academic rigor of a student's course load in high school suggests a promising additional measure to the assessment of a student's level of college readiness. With that in mind, it should also be reiterated that ARI resulted in a larger amount of differential prediction than the SAT for underrepresented minorities and low-income students. Institutions interested in including ARI as part of the college admission process should thoughtfully consider the benefits and limitations that such a measure would add to the admission decision. Additionally, there were many surprising findings that require further attention and research.

First, in contrast to the work by Adelman (1999; 2006) that found that academic rigor was the strongest correlate of earning a bachelor's degree compared to test scores and HSGPA, the current study found that ARI had the lowest correlation with FYGPA out of the three measures. Though FYGPA and graduation are both considered measures of college success, it is possible that different factors influence whether a student earns high grades in college versus whether a student ultimately graduates. Future research should identify and examine potentially influential variables that might explain the current findings. For example, a student's ARI score and his or her likelihood of graduating might be more influenced by motivational factors, whereas academic factors may play a larger role in predicting FYGPA.

Another difference between Adelman's (1999; 2006) and Wiley et al.'s (2010) research concerns the method by which academic rigor was assessed (transcript data versus self-report). In other words, even though Adelman's and Wiley et al.'s conceptualization of academic rigor both took into account the number of courses and level of content into their scoring rubric, Adelman had access to actual transcript data whereas Wiley et al. relied on self-reported course-taking behavior. Self-reported data, particularly course-taking behavior, has been shown to be quite reliable (Schiel & Noble, 1991); however, this could be contributing to the weaker findings reported here. Future research should address this issue by holding constant the method by which academic rigor is measured and examining whether differences in ARI's relationship with FYGPA and graduation persist.

The correlational analyses reveal that ARI not only is indicative of future performance as signified by its positive relationship with FYGPA, but also allows students to showcase their strengths and abilities via a new dimension: the rigor of high school courses completed.

Finally, the current study found that academic rigor did not account for any additional variance in the prediction of FYGPA above and beyond SAT scores and HSGPA, a finding that is also incongruent with previous research (Adelman, 1999; 2006). Future research should examine why this occurred. As mentioned above, possible explanations for this discrepancy include the measure of academic success (FYGPA) employed and/or the method used to assess academic rigor (self-report) in the current study.

The current findings also highlight the challenges that are faced when relying on course titles to determine a student's academic rigor score. For example, an algebra course at one high school may vary widely in content and difficulty than an algebra course offered at another high school; however, the ARI scoring rubric would treat the courses the same. If more fine-grained knowledge of the curriculum associated with a high school course at a specific high school was available, the ARI index could take that information into account, and the validity and utility of the measure would potentially increase. In sum, this study builds on the literature of academic rigor and college success. Specifically, the results presented here reveal that ARI predicts college performance coupled with previous findings of smaller subgroup differences (Wiley et al., 2010), suggesting that the ARI holds promise as a new college readiness measure.

References

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

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., Daniel, B., & Berkovits, I. (2003). *Postsecondary attainment, attendance, curriculum, and performance: Selected results from the NELS:88/2000 Postsecondary Education Transcript Study (PETS), 2000* (NCES 2003-394). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Berkner, L., & Chavez, L. (1997). *Access to postsecondary education for the 1992 high school graduates* (NCES 98-105). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning Test* (College Board Research Report 2000-1). New York: The College Board.

Camara, W., & Kimmel, E. (Eds.) (2005). *Choosing students: Higher education admissions tools for the 21st century*. Mahwah, NJ: Lawrence Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Greene, J. P., & Winters, M. A. (2005). *Public high school graduation and college-readiness rates: 1991-2002* (Education Working Paper No. 8, February 2005). New York: Manhattan Institute for Policy Research. Retrieved from http://www.manhattan-institute.org

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley and Sons.

Horn, L., & Kojaku, L. K. (2001). *High school academic curriculum and the persistence path through college* (NCES 2001-163). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average* (College Board Research Report 2008-5). New York: The College Board.

Lawley, D.N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, 62*(1), 28–30.

Mattern, K. D., Patterson, B. F., Shaw, E. J., Kobrin, J. L., & Barbuti, S. M. (2008). *Differential validity and prediction of the SAT* (College Board Research Report 2008-4). New York: The College Board.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington DC: US Department of Education.

National Governors Association Center for Best Practices & the Council of Chief State School Officers. (2010). *Reaching higher: The common core state standards validation committee.* Washington, DC: National Governors Association.

Patterson, B. F., & Mattern, K. D. (2011). *Validity of the SAT for predicting first-year grades: 2008 SAT validity sample* (College Board Statistical Report 2011-5). New York: The College Board.

Patterson, B. F., Mattern, K. D., & Kobrin, J. L. (2009). *Validity of the SAT for predicting FYGPA: 2007 SAT validity sample* (College Board Statistical Report 2009-1). New York: The College Board.

Pearson, K. (1902). On the mathematical theory of errors of judgment, with special reference to the personal equation. *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character, 198*, 235–99.

Powers, D. E. (2004). Validity of Graduate Record Examinations (GRE) General Test scores for admissions to colleges of veterinary medicine. *Journal of Applied Psychology, 89*, 208–219.

Schiel, J., & Nobel, J. (1991). *Accuracy of self-report course work and grade information on high school sophomores* (ACT Research Report 1991-6). Iowa City: ACT, Inc.

Wiley, A., Wyatt, J., & Camara, W. J. (2010). *The development of a multidimensional index of college readiness for SAT students*. (College Board Research Report 2010-3). New York: The College Board.

Willingham, W. W., & Breland, H. M. (1982). *Personal qualities and college admissions*. New York: College Entrance Examination Board.

Wyatt, J., Wiley, A., Camara, W. J., & Proestler, N. (2012). *The development of an index of academic rigor for college readiness* (College Board Research Report 2011-11). New York: The College Board.

Appendix

Table A1.

Corrected (Observed) Correlation Matrix of SAT, HSGPA, and ARI

	HSGPA	SAT-CR	SAT-M	SAT-W	ARI
HSGPA		.45	.49	.49	.50
SAT-CR	(.21)		.72	.83	.58
SAT-M	(.25)	(.50)		.72	.62
SAT-W	(.25)	(.71)	(.51)		.60
ARI	(.30)	(.37)	(.39)	(.39)	

Notes: N: number of students = 145,131. Pooled within-institution, restriction-of-range corrected correlations are presented above the main diagonal. The observed correlations are shown in parentheses below the main diagonal.

The Research & Development department actively supports the College Board's mission by:

- Providing data-based solutions to important educational problems and questions
- Applying scientific procedures and research to inform our work
- Designing and evaluating improvements to current assessments and developing new assessments as well as educational tools to ensure the highest technical standards
- Analyzing and resolving critical issues for all programs, including AP®, SAT®, PSAT/NMSQT®
- Developing standards and conducting college and career readiness alignment studies
- Publishing findings and presenting our work at key scientific and education conferences
- Generating new knowledge and forward-thinking ideas with a highly trained and credentialed staff

Our work focuses on the following areas

Admission	Measurement
Alignment	Research
Evaluation	Trends
Fairness	Validity

Follow us online: collegeboard.org/research

CollegeBoard

CollegeBoard

12b-5291