Journal of Educational Psychology
2009, Vol. 101, No. 4, 964–981

© 2009 American Psychological Association
0022-0663/09/$12.00 DOI: 10.1037/a0015906

Are SSATs and GPA Enough? A Theory-Based Approach to Predicting Academic Success in Secondary School

Elena L. Grigorenko
Yale University

Linda Jarvin
Tufts University

Ray Diffley III, Julie Goodyear, and Edward J. Shanahan
Choate Rosemary Hall

Robert J. Sternberg
Tufts University

Two studies were carried out to predict academic success in the highly competitive environment of a private preparatory school, Choate Rosemary Hall. The 1st study focused on the question of whether there are indicators beyond middle school grade-point average (GPA) and standardized test scores that might enhance the validity of measures for predicting success of students attending Choate. The results indicated the importance of taking into account aspects of self-regulated learning (SRL), such as academic self-efficacy, academic motivation, academic locus of control, and measures of the WICS (Wisdom, Intelligence, Creativity Synthesized) theoretical framework. Both sets of SRL and WICS indicators demonstrated incremental validity in predicting success at Choate. The 2nd study preliminarily evaluated the value of including indicators of aspects of the SRL and the WICS theoretical framework into the Choate admission process. The results of this study examined the utility of using quantified indicators other than middle-school GPA and standardized test scores for making admission decisions.

Keywords: academic success, secondary school, self-regulated learning (SRL), WICS (Wisdom, Intelligence, Creativity Synthesized)

Standardized tests play a major role in the selection processes for private secondary as well as for tertiary education in the United States. Most of the research literature, however, focuses on tertiary education—on tests taken during the high school years to determine entry and to predict success in college. The most commonly used tests for this purpose are the SAT and the ACT. Numerous studies have demonstrated the usefulness of the SAT and ACT as predictors of college success (e.g., Bridgeman, McCamley-Jenkins, & Ervin, 2000; Noble & Sawyer, 2002). There may be means of improving the predictions provided by these admissions tests, however.

There is preliminary evidence that when these tests are augmented with more wide-ranging measures, the predictive validity of the combined battery can be significantly higher than that of the conventional admissions tests alone (e.g., Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Sternberg & The Rainbow Project Collaborators, 2006). Investigators also have sought to apply modern cognitive theories to augment the prediction of performance at the graduate and professional levels of education (Hedlund, Wilt, Nebel, Ashford, & Sternberg, 2006).

In contrast to this rich literature on college-level tests, there is a much smaller body of literature on the predictive validity of secondary-level standardized tests (i.e., tests taken for admission to high schools, usually private high schools), such as the Secondary School Admission Test (SSAT) and the Independent Schools Entrance Examination (ISEE). The SSAT and ISEE are similar to the SAT in that they measure verbal, quantitative, and reading skills. The SSAT has been found to be predictive in a variety of settings, although the amount of research is limited. Specifically, the SSAT has been found to predict performance in secondary school Latin (Schuerger & Dizney, 1967) and to be useful in the identification of academically talented elementary school students (Lupkowski-Shoplik & Assoline, 1993; Mills & Barnett, 1992). When administered in modified form, it also can be useful with accommodations given to students with learning disabilities (Beattie, Grise, & Algozzine, 1983).

In this article, we investigated the dynamics of secondary-school GPA when predicted by the SSAT or ISEE and a variety of other indicators. These indicators are related to those used in past studies of college GPA, which predicted over and above the SAT and high-school grades (in particular, by Sternberg & the Rainbow Project Collaborators, 2006). The main objectives of the study

Elena L. Grigorenko, Child Study Center, Department of Psychology, and Department of Epidemiology & Public Health, Yale University; Linda Jarvin, Center for the Enhancement of Learning and Teaching, Tufts University; Robert J. Sternberg, Department of Psychology, Tufts University; Ray Diffley, III, Julie Goodyear, and Edward J. Shanahan, Choate Rosemary Hall.

Julie Goodyear is an employee of Foundation for a Greater Opportunity, which is funded by Carl lcahn; her husband is an employee of Choate Rosemary Hall.

This research was supported by the Icahn Family Foundation and Choate Rosemary Hall. The authors express their gratitude to David Francis for his comments on the growth modeling and Robyn Rissman and Mei Tan for their editorial assistance.

Correspondence concerning this article should be addressed to Robert J. Sternberg, Tufts University, Ballou Hall, 3rd Floor, Medford, MA 02155. E-mail: robert.sternberg@tufts.edu

were twofold. First, we wanted to investigate what differentiates students' performance at Choate Rosemary Hall, a highly selective college preparatory school (hereafter, "Choate" or "the school"). Each entering class at Choate is admitted on the basis of the students' middle-school GPAs, SSATs, and teacher recommendations. At admission, new students look quite comparable on these indicators. Nevertheless, the students show dramatic differentiation in performance during their tenure at Choate. Thus, factors other than academic abilities and competencies appear to be at play. For example, although Choate is a school and all of the students have been to school before, because it is primarily a boarding school and most students have not been to boarding schools before, a student's ability to adjust quickly to a new environment is also important. So, one objective was to identify indicators that might account for such differentiation in Choate GPA over time. Second, we wanted to transfer some of the psychological indicators differentiating students' academic performance at Choate to Choate's admission practices. The objective of this part of the work was to see whether students applying to Choate could be differentiated by indicators other than middle-school GPAs and standardized tests. In general, we wished to ascertain whether the selection processes could be enhanced by the use of these indicators. If so, these indicators might not only be instrumental in building the best "suited" classes for Choate but also give the admission office a new, informative, and empowering way to talk about admission with prospective families.

THEORETICAL FRAMEWORK

Two major theoretical contributions form the foundation for this work. A body of theoretical work on self-regulation (Boekaerts, Pintrich, & Zeidner, 2000), especially theories of self-regulated learning (SRL; Pintrich, 2000; Schunk, 2005), contributed to the research presented here. In addition, the framework of the WICS (Wisdom, Intelligence, Creativity Synthesized) theory developed by Sternberg (2003), an expansion of the theory of successful intelligence (Sternberg, 1996), structured the work. Below, we briefly describe key assumptions of these theories.

The SRL Perspective

The SRL perspective is characterized by four general assumptions (Pintrich, 2000). The first two assumptions of this perspective are substance based and the third and fourth are process based. First, students are active participants in the learning process. Thus, no learning process can occur that bypasses the individual characteristics of the learner, and no environment can "make" everyone learn. Second, students have the capacity, at least potentially, to monitor, control, and regulate aspects of their own cognition, motivation, and behavior; students influence, at least partially, the environment in which they learn. Third, students can set standards or goals to aim for in their learning and, while realizing these aims, can monitor their progress and regulate their cognition, motivation, and behavior. Fourth, learning environments (Assumption 1), students' individual characteristics (Assumption 2), and students' goals (Assumption 3) interact dynamically.

Our application of the SRL perspective is also enriched by a variety of other approaches. Specifically, these approaches included Dweck's (1999) theory that learners who believe that they can improve their intelligence perform better in challenging academic environments than do learners who believe their intelligence is fixed. In addition, we used ideas from Bandura's theory of self-efficacy, according to which it helps in accomplishing a task to believe in one's own ability (Bandura, 1996); also relevant are Luthar, Cicchetti, and Becker's theory of resilience, according to which people who demonstrate resilience in the face of failure have better chances of succeeding in the long term (Luthar, Cicchetti, & Becker, 2000); and Rotter's theory of locus of control (Rotter, 1990).

The WICS Theory

WICS builds on Sternberg's earlier models (for a review, see Sternberg, 2003) but differs in that it systematically synthesizes wisdom, intelligence, and creativity. According to WICS, wisdom, intelligence, and creativity synthesized provide a basis for turning out competent and responsible citizens. Such citizens are expected to use (a) creativity to generate new ideas and problems as well as possible solutions to the problems, (b) analytical intelligence to evaluate the quality of these solutions, (c) practical intelligence to implement decisions and persuade others of their value, and (d) wisdom to ensure that these decisions help achieve a common good over the long and short terms. Thus, we argue that a predictive assessment should involve an evaluation of not only cognitive abilities—broadly defined as analytical, practical, and creative—but also elements of responsible reasoning and moral judgment that are integral to wisdom (Sternberg, 2003). This theoretical approach framed our research with regard to what kinds of competencies are essential to student success.

EMPIRICAL FRAMEWORK

The research presented here was structured around the question posed in the title of this article: Are SSATs and middle-school GPA enough for predicting academic success in secondary school, or could and should they be augmented to improve prediction? The SRL and WICS theoretical approaches framed our research with regard to how the competencies delineated earlier needed to be measured. Specifically, we decided to measure relevant competencies through (a) students' performance on competence-related tasks; (b) students' own appraisals of their academic goals, motivation, and their awareness of the learning environment and their academic success in it; and (c) teachers' comments on how the self-regulatory processes of their students change over the period of the freshman year. The work was conducted with multiple samples over a number of years. Each of the studies had a specific objective; each contributed to answering a specific question in addition to the overarching research question. Correspondingly, we present these studies separately and then provide a general discussion of the results.

STUDY SETTING

All research was carried out at Choate (Wallingford, Connecticut). It is a secondary-level boarding school that was initially founded in 1890 as Rosemary Hall (for girls), with Choate School (for boys) following 6 years later. Rosemary Hall was created for the daughters of families who sought a more intellectually stimu-

lating academic program for their daughters than would have been available for girls at home. Choate, like other boarding schools at the time, was created to prepare the sons of well-to-do families to be civic leaders. The schools evolved a great deal through the middle of the century, and then even more so when they merged in 1972. Although transformed, the school continued to include academic excellence and civic responsibility as foundations of their program. However, it has since broadened its outreach to become a geographically, economically, and culturally diverse secondary institution. Choate Rosemary Hall is a school of 840 students, 640 boarding students from 41 states and 33 countries, and 200 day students. One of the top academic boarding schools in the country and one of the most selective, it continues to attract and educate bright and motivated students. With 240 different courses and advanced placement courses in 25 different areas, it provides its students with a rich and challenging academic environment. With 80 interscholastic teams and 60 extracurricular clubs, life outside the classroom is also active. The college admission focus begins early and is competitive. The academic, extracurricular, and social facets of Choate life are advanced. It is a privilege to attend this school for its resources are deep and broad. Choate's admission goal is to ensure that the students it accepts are hungry for these opportunities and that they will adjust to and flourish in the environment. Finding students who will thrive rather than flounder is important for everyone involved (i.e., for students, their families, teachers, and administrators). By providing a challenging, thoughtful, and varied secondary school experience, Choate hopes to prepare its students to be leaders in their future fields and communities.

Study 1

Study 1 comprises a pilot study that included a smaller group of Choate students (hereafter, Pilot Study) and a main study (hereafter, Main Study) that included a larger group of students constituting a whole class. All participation in this research was voluntary.

Pilot Study

The sample in this study comprised high-ability students from lower socioeconomic strata selected to enter the school on scholarship support provided by the Icahn Charitable Foundation and/or the Foundation for a Greater Opportunity. The Icahn Scholars Program identifies motivated and highly able middle-school students from disadvantaged backgrounds and provides 10–18[1] of them each year with a fully funded Choate education. Scholars are identified by teacher and parent recommendations, middle-school GPAs, standardized test scores (SAT, SSAT, and ISEE scores), and a semistructured interview conducted by a representative of the Foundation(s).

The importance of programs such as the Icahn Scholar Program is difficult to overstate. In a global world in which exposure to education and sources of rapidly accumulated information will only continue to be critical to participation, and in which some underresourced populations never move out of a 15-block radius, the benefits of bringing an Icahn Scholar to Choate are many. The child learns and lives in an environment that has crystallized the country's (and to some degree, the world's) highest standards of balanced academic and nonacademic education. The Icahn Scholar is introduced to, and introduces his or her family and often friends and neighbors to, this broader world. Additionally, because the Icahn Scholar Program is not entirely composed of students of color, these students refute the myth that poverty is only a factor for families of color. In turn, the graduates bring with them their Choate experiences as they move on to college and beyond. When one of the first Icahn Scholars graduated and moved on to Harvard, she wrote, "I find I am not homesick. I am Choate-sick."

On another level, the non-Icahn Scholar students at Choate learn about very different home environments, as they live with Icahn Scholars from different backgrounds (a dangerous inner city, a Navajo reservation), and this broadens the horizons of the more resourced students. As a result, when Choate graduates go into the marketplace, they are more informed, more empathetic, and more diversified citizens.

This research originated as a project to determine which qualities were important for the Icahn Scholars' academic success. When expanded to include the entire Choate student body, the results were at least equally valid. Although Choate is unusual in its cost, it is similar to any other high school that prides itself on educating students to participate in competitive educational and labor environments. Although not all high schools may have a preponderance of motivated, bright students, all high schools provide advanced courses for students. As No Child Left Behind continues to stress test scores, all schools will become more achievement driven and more focused on test results and college credits. As schools look to change their students into more academically focused populations, the results from this work can provide information on what skills to teach to increase achievement. This would be universally beneficial to all schools—to schools already educating bright, motivated students and to schools wanting to give their students the skills to become more successful students. Although this work centered on an admission process, the results can be extrapolated into classroom use to enhance learning skills and adaptability to whatever new academic environment needs to be faced throughout one's educational pathway.

Method

Participants

Over the 3 years of the Icahn Scholars Program's existence prior to this study, 55 students were accepted for the program on the basis of their school grades and standardized test scores, parent and teacher recommendations, and results of individual interviews, conducted by one of the authors. However, when admitted, some of the Icahn students did not fare as well academically as the Choate admission officers had hoped; such "unfulfilled hopes" are also true for the student population at Choate, in general. Some students do not do as well in their new school as their admission materials would predict; others, of course, do better than expected, based on their standardized scores. This research was motivated by our desire to understand the causes of this diversification of performance, which occurred among the Icahn Scholars, who,

[1] Numbers vary with available funding.

prior to their arrival at Choate, appeared to be of roughly comparable ability and motivation. Specifically, within the first semester at Choate, the students demonstrated a high level of variation in their Choate GPAs. First, some students ($n = 4$, out of 55 across 3 years of admissions) were not able to complete their studies and left the school. Second, whereas the pre-Choate[2] GPA for the 55 Icahn Scholars admitted prior to the beginning of this research was at a mean value of 3.84 (out of 4) with a standard deviation of 0.30, their postadmission GPAs at Choate varied dramatically. The mean value was 2.89 with a standard deviation of 0.68 in their first trimester at Choate. Thus, the question was whether a battery of assessments could be developed and administered that would be predictive of academic success for the Icahn Scholars in the Choate environment so that the quality of admission decisions could be improved.

The full sample of participants in this study included 51 students, of whom 24 (47%) were girls and 27 (53%) were boys. All were approximately 180 months (15 years) old at the year of admission. The majority came from diverse ethnic groups of color; approximately 35% were White students from low-SES backgrounds.

Procedure

Students were evaluated while in summer school at Choate, prior to the beginning of their first fall semester at Choate. All assessments were group administered and took approximately 40 min to complete. All materials were preprinted at the Psychology of Abilities, Competencies, and Expertise (PACE) Center[3] and administered and scored by the PACE research team. The Icahn Scholars were debriefed on the purposes of the work, which were explained as attempting to improve the selection procedures into the program and to maximize the likelihood of the Icahn Scholars' succeeding at Choate. After contributing their time to this research, the students were offered treats (pizza and/or dessert).

Measures

Self-reports. In developing the materials for this study, three limitations surfaced. First, the group of students was fairly homogeneous with regard to their middle-school GPAs and their standardized test scores. Thus, it was not anticipated that any analytical ability measures would greatly differentiate among them. Second, the power of the sample was fairly low ($n = 51$), limiting the ability to introduce a large number of measures in addition to middle-school GPA and the three SSAT indicators (Verbal, Quantitative, and Reading). Third, the school was interested in enhancing its ability specifically to predict Choate GPAs. As a result of these constraints, a decision was made to focus on self-reported characteristics within the SRL theoretical framework, assessing students' (a) self-awareness of their individual characteristics as learners and (b) capacity to regulate their learning through their academic self-esteem, self-efficacy, locus of control, and motivation. For simplicity, these measures are referred to collectively as PACE measures (PACE Battery). The devised self-report scales were as follows.

The Self-Esteem Scale captured students' perception of their own individual characteristics (e.g., cognitive abilities, personality traits, skills in dealing with peers and adults). An illustration[4] of an item typical of this scale is "You think you are usually intelligent." The scale had nine items, and its internal consistency reliability was acceptable (Cronbach's $\alpha = .85$).

The Academic Self-Efficacy Scale included five items. An illustration of an item typical of this scale is "I can plan my time effectively to get my work done." The Cronbach's alpha for the scale was .75.

The Academic Locus of Control Scale included two items, whereby the students attributed their academic successes and failures either to themselves or to other forces. The Cronbach's alpha was .61.

The Intrinsic/Extrinsic Academic Motivation Scale was designed to grasp students' balance of motivation as driven by their internal forces (e.g., interest) versus external rewards (e.g., grades). An example of an item typical of this scale is "It is important to get the best grades in your class." The scale had 11 items, and its internal consistency was acceptable (Cronbach's $\alpha = .78$).

For all scales, students were asked to rate their responses on a 1–7 scale ranging from 1 (*strongly agree*) to 7 (*strongly disagree*.)

School reports. The school provided (a) demographic data (gender, age, ethnicity, and citizenship); (b) grades for all subjects the students took during their 4 years at Choate, summarized as GPAs for the fall, winter, and spring trimesters for 4 consecutive years (i.e., 12 time points); and (c) preadmission data consisting of students' middle-school GPA, their standardized test results (i.e., three subscores—Verbal, Quantitative, and Reading—of the SSAT).[5]

Results

This study (Pilot Study) was aimed at understanding the contributions of middle-school GPA, SSAT/ISEE scores, and PACE-designed measures to predicting success at Choate, as captured by Choate GPA. Correspondingly, we present the results in three blocks. First, we present descriptive statistics. Second, we present growth analyses of Choate GPA. Third, we present summative regression analyses predicting average GPA while at the school.

Descriptive Statistics

Table 1 presents descriptive statistics for the GPA indices used in this study. As for other relevant indicators, the means and standard deviations (shown in parentheses) for the SSAT indices were 68.0 (24.4), 77.6 (19.9), and 78.2 (16.5) for Verbal, Reading,

[2] This term is used interchangeably with the term *middle-school GPA*.

[3] The PACE (Psychology of Abilities, Competencies, and Expertise) Center was based at Yale University and in 2006 moved to Tufts University. All new measures developed for this work were developed at PACE and, therefore, are referred to collectively as PACE measures/indicators or as the *PACE Battery*.

[4] All assessment items used in this work belong to Choate. Therefore, here we use examples illustrating, but not sampling, the actual items used in this research. For more detail, please contact Raymond Diffley, III, Director of Admission, Choate Rosemary Hall, rdiffley@choate.edu

[5] During the course of this work, the SSAT changed its scaling rules, so the raw scores differ for different studies presented here. Please see http://www.ssat.org

Table 1
Means and Standard Deviations for Indices Used With Icahn Scholars in Study 1, Pilot Study

	GPA			
	Pre-choate		Choate	
Indicator	*M*	*SD*	*M*	*SD*
Time				
Baseline	3.84	0.30		
Freshman year				
Fall			2.78	0.68
Winter			2.91	0.71
Spring			3.00	0.66
Sophomore year				
Fall			3.06	0.58
Winter			3.09	0.62
Spring			3.07	0.64
Junior year				
Fall			3.11	0.60
Winter			3.08	0.63
Spring			3.23	0.65
Senior year				
Fall			3.30	0.51
Winter			3.32	0.59
Spring			3.24	0.64

Note. GPA = grade point average.

and Quantitative, correspondingly. The descriptives for PACE SRL measures were 3.44 (1.19) for self-esteem, 3.35 (0.82) for Academic Motivation, 2.54 (0.95) for Academic Locus of Control, and 2.70 (.092) for Academic Self-Efficacy. For the outcome variable(s) of interest, the 12 indices of Choate GPA, there were no gender differences. There were asystematic ethnicity-related group differences at different time points, when ethnicity was classified into two groups (White and students of color, as captured by the ethnicity status variable). Specifically, White students slightly outperformed students of color (for 7 out of 12 data points, with p values ranging from .048 to .008). The repeated measures analyses of variance indicated the presence of the main effect of ethnicity status, $F(1, 45) = 6.55$, $p < .01$, $\eta^2 = .13$, although there were no interaction effects with the time variable. In addition, average Choate GPA differed for the two groups with regard to their ethnicity status (3.32 vs. 2.82), $F(1, 45) = 7.40$, $p < .01$, with White students outperforming students of color. On the basis of these results, we included the variable ethnicity status but not gender in subsequent analyses.

Conditional Growth in Choate GPA: What Matters?

As is apparent from the data presented in Table 1, Choate GPAs appear to increase over the period of time at school, but this growth is characterized by some fluctuation. To investigate the changes in Choate GPA over time in the group of Icahn Scholars, we fitted two-level growth models, where Level 1 was represented by growth trajectories of Choate GPA, and Level 2 was represented by student-level data, specifically, ethnicity status, pre-Choate GPA, SSAT indicators, and PACE SRL indicators. To complete these analyses, we used hierarchical linear modeling (HLM) software, Version 6 (Raudenbush, Bryk, Cheong, & Congdon, 2000).

First, an unconditional linear growth model was fitted using Icahn GPAs over the 12 trimesters at Choate. The results confirmed the dynamics in GPAs evident from Table 1: There was linear growth in Icahn GPAs over time, with the starting point at approximately 3.05 and a growth rate of .009 GPA per month or .027 GPA per trimester, $T(468) = 6.28$, $p < .001$. Now, what variables among those available in this study can predict either the starting point or the growth rate of Choate GPAs?

To answer this question, we fitted a linear growth model of GPAs at Level 1 and compared whether there were significant differences in rates of growth if the growth were conditioned on the collected indicators (Level 2). All models were fitted separately for separate indicators so that the variables specified at Level 2 were defined as predictors of both the intercept and the slope of growth. The time at Choate was captured by academic months at Choate (with summer months not counted). Table 2 presents the outcome of these analyses.

A number of observations can be drawn from Table 2. The models differed in terms of the impact of specific variables on the starting GPA value and GPA growth rate. Specifically, the intercept was influenced by (a) ethnic status, with students of color starting at a lower GPA than did White students ($p < .01$), with students of color approximately .5 GPA points behind, (b) middle-school GPA, so that students with higher pre-Choate GPA started at higher GPA at Choate ($p < .001$), with one unit of pre-Choate GPA, resulting in approximately.9966 units of Choate GPA; (c) SSAT Quantitative (SSAT-Q; $p < .05$), with one SSAT-Q unit predicting .0121 units of Choate GPA; (d) SSAT Reading (SSAT-R; $p < .05$), with one SSAT-R unit predicting .0083 units of Choate GPA; and (e) Academic Self-Efficacy ($p < .005$), with a unit of stronger efficacy predicting .2909 units of Choate GPA. Only three variables affect the rate of growth in GPA: ethnicity status ($p < .05$), with students of minority background showing slower growth; SSAT Verbal ($p < .001$), with students with higher SSAT scores showing slower growth rates; and Academic Locus of Control ($p < .001$), with students with an external locus of control exhibiting a higher rate of GPA growth.

To summarize these results in a single model, we fitted a complex equation in which Choate GPA intercept and slope were predicted by those variables whose contributions were statistically significant in single-predictor models (see Table 2). Fitting such a combined model with multiple variables resulted in selected variables losing their significance. Dropping these statistically insignificant variables demonstrates no loss of fit, (χ^2difference with 5 *df* was 5.46, *ns*). Therefore, here we discuss only the parameters from that final reduced model. According to that model, two variables influence the GPA intercept (starting value). These variables were middle-school GPA ($p < .005$), with one unit of pre-Choate GPA predicting .8022 units of Choate GPA, and Academic Self-Efficacy ($p < .01$), with one unit of self-efficacy predicting .2153 units of Choate GPA. The rate of growth was predicted by three variables, ethnicity status ($p < .005$), with minority students gaining GPA at a lower rate; SSAT Verbal ($p < .001$), with higher scores associated with slower growth in GPA; and Academic Locus of Control ($p < .05$), with more externally oriented students exhibiting higher rates of growth.

Table 2
Change in Choate Grade Point Average (GPA; Study 1, Pilot Study)

Model	Intercept (starting value)			Slope (growth rate)		
	Coefficient	*T*	*p*	Coefficient	*T*	*p*
Single-predictor model						
Parameter model						
Ethnicity status	−.4681	−2.70	.010	−.0063	−2.22	.026
Pre-Choate GPA	.9966	3.77	<.001	−.0034	−0.71	.481
SSAT Verbal	.0063	1.73	.091	−.0002	−4.06	<.001
SSAT Quantitative	.0121	2.27	.029	−.0000	−0.55	.580
SSAT Reading	.0083	2.18	.035	−.0001	−1.12	.269
Self-esteem	−.0786	−0.94	.354	.0004	0.34	.738
Academic Self-Efficacy	−.2909	36.97	.003	.0008	0.50	.619
Academic Locus of Control	−.0729	−0.74	.462	.0048	3.37	.001
Academic Motivation	−.1124	−0.97	.339	−.0004	−0.24	.813
Multiple-predictor (combined) model						
Full model						
Ethnicity status	−.0423	−0.26	.800	−.0090	−3.19	.002
Pre-Choate GPA	.6729	2.50	.017			
SSAT Verbal				−.0002	−3.79	<.001
SSAT Quantitative	.0086	1.89	.066			
SSAT Reading	.0036	0.92	.361			
Academic Self-Efficacy	−.1911	−2.23	.031			
Academic Locus of Control				.0030	2.04	.041
Reduced model						
Ethnicity status				−.0090	−3.16	.002
Pre-Choate GPA	.8022	3.13	.004			
SSAT Verbal				−.0002	−3.78	<.001
Academic Self-Efficacy	−.2153	−2.56	.015			
Academic Locus of Control				.0030	2.03	.042

Note. SSAT = Secondary School Admission Test.

Predicting Averaged Choate GPA

The growth model analyses demonstrated the importance of indicators from all types of variables considered in this work—previous GPA, standardized tests, and indicators of SRL. Some of these variables were important predictors of the starting point at Choate, whereas others predicted the dynamics of growth in GPA. In the analyses that follow, we attempted to predict the averaged Choate GPA. Such averaged GPAs are used for evaluation of college applications and, subsequently, for predicting college GPA.

The results of these analyses are shown in Table 3. Two regression equations were fit.[6] The first regression included only the SSAT and the PACE SRL indicators (see the top portion of Table 3). Collectively, SSAT indicators explained about 15% of the variance in Choate GPA, but not a single standardized coefficient was statistically significant (the model's p value was .07). The introduction of PACE SRL indicators explained approximately an additional 17% of the variance in school GPA, generating a final R^2 of .32, with $p < .05$ for the model. The second regression included middle-school GPA (see the bottom portion of Table 3). Collectively, SSAT indicators and middle-school GPA explained approximately 34% of the variance (model $p < .005$). The addition of the PACE SRL indicators added about 8% more of the explained variance, bringing the squared correlation to .42 (model $p < .008$). When the reduced regression was fit, which included only the statistically significant coefficients from the analyses presented earlier (i.e., pre-Choate GPA and Academic Self-Efficacy), the model explained 34% of the variance ($p < .001$), of which 24.6% was attributed to middle-school GPA, and 9.6% was attributed to Academic Self-Efficacy. Both coefficients were statistically significant ($\beta = .41$, $p < .01$ and $\beta = -.32$, $p < .05$, for pre-Choate GPA and Academic Self-Efficacy, respectively, with higher pre-Choate GPA and higher Academic Self-Efficacy predicting higher Choate GPA).

Discussion

The results of this study suggest that middle-school GPA and standardized test results, although informative in predicting high-school GPA, predict, at least in this sample, only a portion of the variance in Choate GPA. Additional indicators capturing characteristics of SRL are informative not only in increasing the prediction of the absolute values of GPA but also in predicting the rate of growth in GPA across the 12 trimesters (or 36 academic months) at Choate.

Although interesting, these observations are preliminary because of the small size and consequent lack of generalizability of the studied sample of Icahn scholars. In addition, the repertoire of

[6] Of note is that the inclusion of the ethnicity status variable did not improve the R^2. Because the obtained coefficient was not significant, this variable was omitted from analyses.

Table 3
Incremental Prediction of High School GPA of Icahn Scholars Using SRL-Related Measures (A) Above and Beyond SSAT and (B) Above and Beyond SSAT and Pre-Choate GPA (Study 1, Pilot Study)

Measure	Step 1	Step 2
	A	
SSAT		
Verbal	.084	.043
Quantitative	.261	.267
Reading	.183	.117
PACE SRL indicators		
Self-esteem		.229
Academic Self-Efficacy		−.519**
Academic Locus of Control		−.025
Academic Motivation		.019
R^2	.159	.324
	B	
SSAT		
Verbal	.024	.003
Quantitative	.182	.205
Reading	.198	.144
Pre-Choate GPA	.437**	.340*
PACE SRL indicators		
Self-esteem		.146
Academic Self-Efficacy		−.379*
Academic Locus of Control		−.009
Academic Motivation		.006
R^2	.339	.421

Note. Entries are standardized beta coefficients. GPA = grade point average; SRL = self-regulated learning; SSAT = Secondary School Admission Test; PACE = Psychology of Abilities, Competencies, and Expertise.
* $p < .05$. ** $p < .01$.

measurements was limited to self-reports only. Yet, this Pilot Study formed the foundation for our continuing the work.

Main Study

The main objective of this study was to capitalize on the preliminary findings from the Pilot Study and overcome its three main limitations by recruiting a (a) larger and (b) more representative group of participants, and by (c) including not only self-reports but also performance-based measures. In the Main Study, we worked with a whole freshman class of newly admitted Choate students.

Method

Three parameters of this study are essential in the description of its methodology. First, all of the assessments are theory driven (see the introduction). Second, to maximize the information gained from these assessments, the assessments were administered to an entire class of freshman students while they were adjusting to the environment at Choate; thus, a multiple time point longitudinal investigation of their adjustment to new learning challenges was carried out. Third, a multitrait/multimethod assessment approach was used in which different traits were assessed, different informants (students, teachers, and the school itself) were engaged, and different methodologies (self-report, rating scales, and maximum performance assessments) were used.

Participants

All freshman boys and girls ($N = 152$, 76 boys and 76 girls; mean age = 176.8 months, $SD = 5.3$) admitted to the school for the 2005–2006 academic year were asked to participate in this study. There were 95 (62.5%) White students; the rest ($n = 57$, 37.5%) were students of color.

In addition, 20 teachers employed by the school full time were asked to evaluate the students' adaptation to the school's environment using a structured evaluation form. Among the participating teachers, there were 8 women and 12 men (ranging from 25 to 72 years of age, mean age = 48.1) who had been teaching at the school anywhere from 2 to 50 years (mean length of teaching at the school = 14.95 years); the majority of teachers were White.

Procedure

Students were evaluated three times: in early December ($n = 149$), late January ($n = 152$), and mid-April ($n = 138$) of their freshman year. All assessments were group administered and took approximately 90 min to complete. All materials were preprinted at the PACE Center and administered and scored by the PACE research team. After contributing their time to this research, the students were offered treats (pizza and/or dessert).

Teachers evaluated students five times, with the following due dates for the evaluation forms: early December, late January, early March, late April, and late May of their freshman year. All evaluations were completed by the teachers individually, at their own pace and in locations of their choosing.

Measures

Eight different assessments were administered to the students, and one survey was administered to the teachers; these assessments, collectively, were referred to as the PACE Battery. Some components of the battery included improved versions[7] of the self-reports from the Pilot Study; others were newly developed. As in the Pilot Study, relevant information was obtained through students' school records.

WICS constructs. Driven by the WICS theory, the following constructs were evaluated:

1. Analytical competence was assessed primarily through students' preadmission SSAT scores and preadmission grades (pre-Choate GPA).

2. Practical competence was assessed primarily through an assessment of tacit knowledge of the school's environment (the School Life Inventory) and a set of practical reasoning tasks.

[7] Specifically, we expanded the number of items for each scale, adding new statements while preserving the internal consistency of the scales.

3. Creative competence was assessed primarily through a creative writing task and by an assessment of the ability to deal with novelty (involving selective encoding and recombining of information while applying scientific knowledge to solving problems, i.e., the scientific reasoning task).

4. Ethical reasoning and wise reasoning were assessed using scenarios presenting ethical dilemmas in which students were asked to analyze an ethically charged situation, find a solution to the situation, attribute blame for a particular event, and state their reasons.

These tasks were not designed as pure measures of particular competencies. For example, the scientific reasoning task called for analytical competence when the problem needed to be identified from the description and a proper element of scientific knowledge had to be evoked. Yet, it called for creative competence when the information needed to be selectively encoded from the mass of information presented and the novelty of the problem needed to be addressed so that the problem could be reformulated and specific knowledge invoked to solve it.

SRL constructs. In addition, driven by the SRL perspective, students' self- and teacher-based comments on their academic goal orientation, motivation, and appraisal were recorded. These items were devised to capture students' self-perceptions at different stages of learning, namely, the planning of learning tasks, the monitoring and controlling of performance, and reflection on their accomplishments.

Self-reports. Self-reports were extended and improved versions of the scales were used in the Pilot Study (see above).

The Academic Self-Efficacy Scale included 12 items, on each of which students rated themselves on a scale ranging from 1 (*strongly disagree*) to 9 (*strongly agree*). Items contained statements regarding students' views of their own capabilities to produce desired levels of performance (e.g., "I can cope with any homework assignment"). Cronbach's αs were .77, .81, and .80 for Times 1, 2, and 3, respectively.

The Academic Locus of Control Scale included 16 items, on each of which students rated themselves, or rather their views of the source of control over their academic success at Choate, on a scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). An example of an item typical of this scale is "I can control my grades at school." Cronbach's αs were .74, .76, and .71 for Times 1, 2, and 3, respectively.

The Intrinsic/Extrinsic Academic Motivation Scale included 12 items, on each of which students rated themselves on a scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The items were designed to capture the students' perception of the origin of academic motivation, which can be placed internally (i.e., originating within the student, intrinsically) or externally (i.e., originating from the students' context, extrinsically). Cronbach's αs were .79, .83, and .82 for Times 1, 2, and 3, respectively.

Performance tasks. All tasks were new and developed for the purposes of this research.

The School Life Inventory included 35 items administered in two versions that capture different aspects of the implicit rules of Choate life. The items presented a general scenario characteristic of various aspects of life at Choate (e.g., working on one's homework assignments while there is a birthday party in the dorm) and various choices a student can make in such situations (e.g., stop working on the homework right away and join the party, finish the homework and then join the party, go to the party for a few minutes and then leave to finish the homework, skip the party, and so forth). Each item included seven different behavioral choices; students were asked to rate every choice on a scale ranging from 1 (*not a very good choice*) to 7 (*a very good choice*). Both versions shared 15 identical items and differed in 10 items. Thus, each student received 25 vignettes. These items were scored in two different ways so that the scores assigned to each participant were derived by calculating the squared Euclidean distance (d^2) and the Mahalanobis distance (D^2; the standardized Euclidean distance, where not only absolute distances but also their dispersions are taken into account) of the participant's ratings for each possible solution strategy from the mean ratings of the total sample.

For each of the seven possible solution strategies accompanying each of the 25 vignettes, the sample's mean rating was subtracted from the participant's rating. These computations resulted in a vector of seven simple difference scores for each participant for each of the 25 vignettes presented in one of the versions, and thus $25 \times N$ vectors in all. For the Euclidean distance, the "ordinary" distance between the seven points of the individual's profile and the mean profile of the sample were calculated. For the Mahalanobis distance, the vectors of different scores were each multiplied by the variance–covariance matrix of the seven possible response strategies from which the difference scores were created. The resulting 7×1 vector was then multiplied by the transpose of the original difference-score vector, resulting in a scalar, D^2. These computations, then, resulted in 25 d^2 and D^2 values per individual, one per vignette, and thus $25 \times N$ in all. The individual's total score for the School Life Inventory was determined by averaging the resulting vignette-level values (25 vignettes for each participant).

Although there is a conceptual difference between using the Euclidean distance measure (d^2) and the Mahalanobis distance measure (D^2), where the former operates only with absolute differences and the latter also includes variation in these distances, the Pearson correlations between d^2 and D^2 were .97, .97, and .96 for Times 1–3, respectively. As noted earlier, scores on the practical ability performance measures were determined with reference to the average, or consensus responses of the sample. Important concerns arise when consensual scoring techniques become imbalanced with regard to race, ethnicity, or gender, as such imbalances might be biased against minority group members; other problems arise with regard to defending the basis of any particular individual's score against the average responses of the sample. However, using an "expert group" as a reference instead of the average responses of the sample might lead to similar problems, for example, with determining the demographic characteristics of those individuals comprising such an expert group. Legree (1995) demonstrated that the ratings of experts and nonexperts on a situational judgment inventory are highly correlated, indicating that a fairly knowledgeable nonexpert consensus can be as sensitive to relative differences in solution quality as an expert consensus. Mayer and colleagues (Mayer, Salovey, Caruso, & Sitarenios, 2003) have shown that an expert panel shows more within-group consistency than does a general sample in selecting the "correct" answer on emotional intelligence items; however, there appears to be a great

deal of between-group agreement in terms of these items, suggesting that both expert panels and general samples tend to agree on the overall correct answers to emotional intelligence items. For the two versions, Cronbach's αs for d^2 were .91 and .93 (Time 1), .95 and .95 (Time 2), and .94 and .94 (Time 3). For D^2, Cronbach's αs were .84 and .83 (Time 1), .92 and .95 (Time 2), and .90 and .93 (Time 3). Given the similarity of the scores obtained by these scoring systems, only the results obtained by calculating the Mahalanobis distance (D^2) are presented here, where not only absolute distances between an individual's and the mean profile of choice are taken into account but also the corresponding variance. Notice that in this task, the higher the score, the greater the deviation from the mean profile of answers, and thus the worse the score.

The practical reasoning in writing assessment contained eight prompts presenting different everyday situations comparable to those experienced by the school's students. For example, students could be presented with the situation of having an assignment due the next day when they had forgotten about the assignment and the deadline. Participants were asked to continue the story by identifying with the main character and developing the next step in the plot. The written products were subsequently reviewed by two independent raters (the same pair of raters for Times 1 and 2 and a different pair—with one continuous rater and one new rater—for Time 3) using scoring rubrics. The corresponding median κs between the two raters were .95, .96, and .93 for Times 1, 2, and 3, respectively.

The ratings were done on three dimensions per item: (a) quality of writing (grammar, spelling, and so on); (b) practicality of the proposed solution; and (c) quality of argument delivered in the excerpt. To process the ratings, the data were subjected to Rasch analysis using Facets (2009); three combined scores per participant were generated per time wave. The reliability estimates for these scores were .83, .61, and .72 (Time 1); .70, .64, and .74 (Time 2); and .73, .66, and .79 (Time 3) for quality of writing, practicality, quality of argument, respectively. These reliability indices indicate the consistency with which the same group of people will score on a different set of items of comparable difficulty and discrimination. The median (across quality of writing, practicality, and quality of argument) reliability estimates for the items (i.e., consistency of responses across the eight practical reasoning tasks) across the three waves of data collection were .96, .95, and .76 for Times 1–3, respectively. These indices suggest that these tasks would behave with similar consistency as a scale when administered to a different group of students.

The creative writing task asked for a brief story under one of five proposed titles: (a) "Too Much, Too Fast"; (b) "The Landing on the Planet Vespa"; (c) "Third Time's the Charm"; (d) "The Spy Was Not Captured After All"; and (e) "When the Music Stopped." Across the three waves, Titles 1–4 were chosen with almost equal frequency (20%–25%); Title 5 was chosen by 8%–11% of participants. Only 4 participants chose the same titles over the three data collection waves (Titles 2 and 5, 2 participants for each title), but a number of participants selected the same title twice (11, 2, 6, 2, and 8 selections for Titles 1–5, respectively).

A number of participants provided answers that were unscorable, that is, that did not correspond to the title/instruction or were not legible (*n*s = 7, 11, and 16 for Times 1–3, respectively). Only 1 participant did not generate ratable answers for any of the three waves. Two independent evaluators, using a scale of 1–6 (low to high), rated the story on two dimensions: (a) quality of written expression and (b) creativity. The corresponding average κs were .33, .72, and .82 for Times 1–3, respectively.

To process the ratings, the data were subjected to Rasch analysis using Facets; a combined score (a synthesis of quality of written expression and creativity) was generated per time wave. Because five different titles were used in this task, three facets were introduced to score these data: title, rater, and participant. The participant-based reliability estimates for these combined scores were .89, .86, and .84 for Times 1–3, respectively. These reliability indices indicate the consistency with which the same group of people will score on a different set of items of comparable difficulty and discrimination. The reliability estimates for quality of written expression and creativity were .98, .88, and .69 for Times 1–3, respectively. Similarly, the five different titles showed acceptable levels of reliability: .71, .83, and .85 for Times 1–3, respectively, indicating that these (or a comparable set of) titles have the potential to generate reliable data when administered to a different sample of participants.

The scientific reasoning task included 15 different word problems describing various situations related to the application of scientific knowledge; students were asked to find a solution using some knowledge of the sciences. Answers were subsequently reviewed, using scoring rubrics, by two independent raters (the same pair) for Times 1 and 2 and by three independent raters (all but one different from Times 1 and 2) for Time 3. Each item was scored on a 4-point scale ranging from 0 (*completely wrong*) to 4 (*completely right*). The corresponding median κs were .92, .94, and .85 (calculated across the three raters) for Times 1, 2, and 3, respectively. The raters' ratings were subjected to Rasch analysis with Facets. For this scale, only one score per participant was acquired, and corresponding rater reliability estimates were .57, .76, and .85 for Times 1–3, respectively. These reliability indices indicate the consistency with which the same group of people will score on a different set of items of comparable difficulty and discrimination. The reliability estimates for the items (i.e., consistency of responses across the 15 scientific reasoning tasks) across the three waves of data collection were .96, .96, and .97 for Times 1–3, respectively. These indices suggest that these tasks would behave with similar consistency as a scale when administered to a different group of students.

The self vs. other(s) attribution of blame assessment included eight brief stories with six possible answer choices. Students were expected to circle only one answer per story, with each choice having a specific expert-determined weight: 0, 1, or 2. Higher values indicate the more expert-valued choices of answers. These scores were generated using a rubric developed by three professional psychologists based on modern literature on moral reasoning. The rubric was then used in scoring by two raters, who were trained prior to scoring to achieve a consensus for at least 80% of the ratings. The scale assessed the student's capacity to determine whether a situation is attributable to his or her own doing, could be someone else's fault, or is viewed as a result of "fate." Cronbach's αs were, disappointingly, only .25, .31, and .48 for Times 1, 2, and 3, respectively.

Teacher reports. Teachers were asked to rate students, using a scale from 1 to 7, on 23 items subgrouped into three categories: self-perception and coping skills (SP&CS, seven items), academic

skills (AS, eight items), and social and practical skills (S&PS, eight items). Here are some illustrative items: "This student can cope with failure"; "Academic skills of this student are strong"; "This student has good relationships with peers," for the three categories, respectively. Twenty teachers from two academic domains—7 teaching various levels of English and 13 teaching various levels of mathematics—were asked to provide five sets of evaluations spread out evenly over time from late November to the end of the students' first year at Choate. Each student was evaluated by two teachers, one in English and the other in math. Teachers were not asked for evaluations earlier in the academic year because they felt that such evaluations would be inaccurate or, at best, less accurate, as they did not know their students as well. In these evaluations, a reversed Likert scale was used, with 1 indicating the strongest presence of a positive characteristic. Because there were so many different raters involved for the overall group, and two raters per student, the data were processed with Facets; three continuous indicators were generated (SP&CS, AS, and S&PS) for all five waves of evaluation. Table 4 summarizes the reliabilities for students, raters, and items for Times 1–5.

School reports. The school provided (a) demographic data (gender, age, ethnicity); (b) grades for all subjects the students took during their first and second years at Choate, summarized as GPAs for the fall, winter, and spring trimesters for 2 consecutive years; (c) information on financial aid (whether the student received any aid or not); and (d) preadmission data consisting of students' GPAs in their previous middle schools and their standardized test results (i.e., three subscores: SSAT Verbal, Quantitative, and Reading).

Results

Findings from this study are presented in three blocks. First, descriptive statistics for all waves of data collection are presented. Then, changes over the first year of education at the school across multiple assessment times are considered in both student and teacher reports. In addition, the changes in Choate GPA over six trimesters are considered. Finally, the power of indicators collected prior to admission by the school and while at the school as predictors of academic achievement (i.e., average Choate GPA) is considered.

Descriptive Statistics

Table 5 presents the descriptive statistics for the various indicators used in this study. Similar to the procedures in the Pilot Study, we screened the outcome variables of interest (GPAs from six trimesters and the GPA averaged across the six trimesters) for group differences on the basis of gender and ethnicity status variables. None of the outcome variables demonstrated group differences, so we carried out all subsequent analyses for the whole sample without stratification by gender or ethnicity.

Changes Over the First Year in Independent Variables and Over the Six Trimesters in Choate GPA

The next set of analyses addresses the question posed by the SRL perspective: We wanted to investigate whether we could detect any changes in the indicators of Choate GPA and competencies in SRL across the first year in the new schooling environment. Specifically, were there any fluctuations in these students' GPAs over their time at Choate? Also, were there any changes across the full first year at the new school in the students' self-reports and performance indicators, and the teachers' evaluations of their students? And, finally, were the changes (if any) in the independent variables related to the changes in Choate GPA?

To analyze these changes, using HLM 6.0, we first fitted an unconditional linear growth model for Choate GPA and all independent variables measured multiple times. We started by fitting an unconditional linear growth model using the GPAs over six trimesters (freshman and sophomore years) at Choate. The results fitted the pattern of GPAs apparent from Table 5: There was linear growth in GPAs over time, with the starting point at approximately 3.14 and a growth rate of .014 GPA per month or .040 GPA per trimester, $T(871) = 8.78$, $p < .001$. These results are similar to those obtained in the Pilot Study, although both the starting point and the average growth of GPAs in this Choate class as a whole appears to be greater compared with these parameters among the three Icahn classes. Given the goal of the Icahn Scholars Program to accept students to Choate who would be at risk of academic failure were they to stay home, this result seems appropriate.

We carried out similar analyses for all PACE indicators on which we had multiple measurements; the *time scale* was defined as months at Choate.

Table 4
Reliability Indicators of Teacher Surveys (Study 1, Main Study)

Indicator	Reliability	Evaluation waves				
		Time 1	Time 2	Time 3	Time 4	Time 5
Self-perception & coping skills	Students	.87	.88	.89	.91	.90
	Raters	.91	.86	.95	.92	.95
	Items	.93	.87	.80	.81	.84
Academic skills	Students	.90	.91	.89	.95	.92
	Raters	.95	.94	.95	.95	.95
	Items	.98	.97	.80	.96	.95
Social & practical skills	Students	.75	.81	.83	.86	.77
	Raters	.91	.94	.93	.93	.94
	Items	.97	.98	.95	.95	.98

Table 5
Means and Standard Deviations for Indices Used With Choate Students in Study 1, Main Study

Assessment	December		January		March		April		May	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Students' self-report indicators										
Academic Self-Efficacy	7.00	0.90	6.86	0.99			6.99	1.04		
Academic Locus of Control	3.53	0.47	3.49	0.57			3.47	0.51		
Academic Motivation	3.86	0.54	3.89	0.64			3.86	0.63		
Students' performance indicators										
School Life Inventory	6.92	3.04	6.97	3.16			6.96	2.90		
Practical reasoning in writing										
Quality of writing	−0.22	1.88	−0.75	1.75			−0.57	1.77		
Practicality	0.00	0.84	0.00	0.88			0.00	0.90		
Quality of argument	0.00	1.07	0.00	1.13			0.00	1.47		
Creative writing	−0.93	3.76	−2.44	4.44			−5.60	5.44		
Scientific reasoning	0.00	0.51	0.00	0.63			0.00	0.81		
Attribution of blame	0.99	0.30	1.09	0.31			0.97	0.37		
Teacher survey[a]										
Self-Perception & Coping	−1.05	0.91	−1.29	1.03	−1.63	1.40	−2.18	1.40	−2.13	1.46
Academic Skills	−1.62	1.00	−1.84	1.15	−1.92	1.78	−2.18	1.75	−2.21	1.74
Social & Practical Skills	−0.99	0.62	−1.73	0.78	−1.57	1.24	−2.02	1.10	−2.28	1.12
GPA										
Pre-Choate	3.78	0.33								
	Fall				Winter				Spring	
Choate										
Freshman year	3.20	0.43			3.20	0.51			3.20	0.52
Sophomore year	3.39	0.56			3.40	0.62			3.37	0.61
SSAT										
Verbal	320.49	14.18								
Quantitative	325.99	13.88								
Reading	313.53	13.89								

Note. GPA = grade point average; SSAT = Secondary School Admission Test.
[a] The scale was reversed, with 1 indicating the highest and 7 indicating the lowest rating.

Overall, there was little change in students' indicators of SRL. Specifically, the unconditional model showed no growth in Academic Self-Efficacy across the three time points. The unconditional model showed a tendency for decline in the internal Academic Locus of Control across the three time points ($B = -0.013$, $SE = 0.01$), $T(437) = -1.74$, $p < .1$, indicating that, with the passing of time, students' orientation toward academics tended to diminish, although this change was not statistically significant. There was no change detected in Academic Motivation over time with an unconditional growth model.

Similarly, we detected little or no change in students' performance measures. The unconditional model of the School Life Inventory detected no change in the students' tacit knowledge of the school's environment across the period from December to May of the freshman year. The practical reasoning task generated three indicators, as described earlier: (a) quality of writing (grammar, spelling, and so on); (b) practicality of the proposed solution; and (c) quality of argument delivered in the excerpt. Correspondingly, three different linear growth models were fitted in these data. The unconditional model fitted into the data for quality of writing indicated a decrease in this indicator over time at the rate of about .08 per Choate academic month ($B = -0.08$, $SE = .03$), $T(437) = -2.52$, $p < .01$. The unconditional model for the indicator of practicality did not show any change in performance over time. Fitting the unconditional model into the data for the quality of argument indicated that there was no change over three assessment times in this variable either. The creative writing unconditional model indicated a substantial change in performance over the three waves of data collection, with the quality of writing getting worse at approximately .8 of the unit of judgment per Choate academic month ($B = -0.81$, $SE = 0.09$), $T(437) = -9.34$, $p < .001$. The unconditional model of scientific reasoning and attribution of blame did not register any changes over time in the performance on these tasks. Further research is needed to understand the mechanisms of these changes (or lack of such); they may be connected to deep motivational structures or to simple exhaustion from accepting and managing repeated tasks over time.

This set of analyses was completed with growth modeling of teacher surveys. As indicated earlier, teacher surveys were scored to generate the following indicators: (a) SP&CS, (b) AS,

(c) and S&PS. The scoring was reversed, with 1 indicating the highest and 7 indicating the lowest rating. Unconditional models fitted for the ratings of students' skills captured a substantial change in these ratings over time in all three domains of teacher ratings ($B = -0.23$, $SE = 0.01$), $T(747) = -17.69$, $p < .001$; ($B = -0.11$, $SE = 0.01$), $T(747) = -8.18$, $p < .001$; ($B = -0.21$, $SE = 0.01$), $T(747) = -16.61$, $p < .001$), for SP&CS, AS, and S&PS, respectively. These findings suggest that teachers' perceptions of students' competencies increased substantially in value over the freshman year. This increase appears to happen at about .2 of the unit of judgment per Choate academic month.

Summarizing these results, three observations need to be made. First, similar to the results in the Pilot Study, Choate GPA showed a significant fluctuation depending on the period of time spent at the school. Second, only one of the self-report scales, Academic Locus of Control, showed fluctuations over time. This result is somewhat contrary to expectations, based on the SRL perspective, and deserves further investigation. Third, all dimensions of the teacher ratings improved with time; thus, it appears that teachers increase their appreciation of the school's students as they get to know them better.

These data provide us with a rather unique opportunity to consider the time-based dynamics in Choate GPA in conjunction with time-based dynamics in the PACE Battery indicators and teacher ratings. Using the software Mplus Version 3 (Múthen & Múthen, 2005), we carried out a set of analyses in which two processes were considered simultaneously, which estimated the associations between the intercepts and slopes of these two processes.

When we analyzed changes in Choate GPAs in conjunction with fluctuations in the indicators from the PACE Battery, a number of significant findings were established; the mean values of Choate GPA were associated with both SRL self-ratings and performance tasks. Specifically, higher mean Choate GPA was associated with higher levels of Self-Efficacy ($B = 0.10$, $SE = 0.04$; $T = 2.68$, $p < .01$), and higher levels of Academic Motivation ($B = 0.05$, $SE = 0.02$; $T = 2.74$, $p < .01$). Similarly, higher Choate GPAs were associated with a number of performance task indicators: better tacit knowledge of Choate ($B = -0.37$, $SE = 0.11$; $T = -3.38$, $p < .001$, for the School Life Inventory); higher performance on the practical reasoning task ($B = 0.11$, $SE = 0.06$; $T = 1.88$, $p < .05$, for quality of writing); higher creativity in writing ($B = 0.36$, $SE = 0.14$, $T = 2.50$, $p < .05$, for creative writing); higher levels of scientific reasoning ($B = 0.06$, $SE = 0.02$, $T = 2.84$, $p < .05$); and more adequate attribution of blame ($B = 0.03$, $SE = 0.01$, $T = 2.85$, $p < .01$).

Only two estimates of the rate of change of the indicators from the PACE Battery were associated with the mean Choate GPA: the Practicality aspect of the practical reasoning task ($B = 0.07$), $SE = 0.03$, $T = 2.16$, $p < .05$), and the scientific reasoning task ($B = 0.07$, $SE = 0.03$, $T = 2.09$, $p < .05$). None of the indicators from the PACE Battery predicted the rate of growth in Choate GPA.

The facets of teacher ratings also showed numerous associations with time fluctuations in Choate GPA (see Table 6 for details). Specifically, dynamics of teacher evaluations of the students on all three dimensions suggest an association of the mean value of Choate GPA (the higher the GPA, the higher the evaluations).

Table 6
Parallel Changes in Teacher Ratings and Choate GPA (Study 1, Main Study)

Teacher ratings	Choate GPA	
	Intercept (B, T)	Slope (B, T)
Self-Perception and Coping Skills		
Intercept	−.11, −3.12**	
Slope	−.26, −7.44***	−.001, −2.42*
Academic Skills		
Intercept	−.21, −5.66***	
Slope	−.32, −8.51***	−.001, −2.41*
Social and Practical Skills		
Intercept	−.07, −1.97*	
Slope	−.17, −4.90**	

Note. GPA = grade point average; B(s) indicate estimated parameters; T(s) indicate corresponding *T* ratios.
* $p < .05$. ** $p < .01$. *** $p < .001$.

In addition, it is apparent that the mean value of Choate GPA was reflected by the rate of growth in the teacher evaluations (for all three variables: SP&CS, AS, and S&PC). Finally, the rate of growth of Choate GPA is associated with the rate of growth in SP&CS and AS. In summary, these results suggest that changes in Choate GPA and changes in teacher ratings are parallel and that it is the dynamics of GPA that are associated with the ratings (i.e., teacher perception of students changing depending on how successful they are at Choate).

The results of these analyses allowed us to make two general observations. First, the mean GPA at Choate appears to be related to a number of psychological constructs captured by the PACE Battery. However, the causal links between these associations are difficult to establish at this point, as a result of the design of the study, the relatively low power of our sample, and, possibly, the influences of other factors. Second, the mean GPA and teacher ratings also appear to be associated; in fact, these associations are strong, and their patterns allow us to hypothesize the causal impact of GPAs on teachers' perceptions of their students.

Predicting First-Year GPA On the Basis of Pre- and In-School Assessments

Up to this point, we have shown that, in the Main Study, (a) we collected reliable data on the newly developed indicators, and (b) these indicators are relevant to predicting the mean values of Choate GPA. Thus, we have already demonstrated the value of the PACE Battery in capturing the average value of GPA at Choate. What we have not yet demonstrated is the relative value of the PACE indicators when compared with the predictive validity of pre-Choate GPA and the three SSAT indicators.

In this section of analyses, we address the specific question of whether students' GPA(s) across the six trimesters of their freshman and sophomore years at Choate can be predicted incrementally, in terms of their absolute value and rate of change, by any of the newly developed measures over the indicators of SSAT and pre-Choate GPA.

To address these questions, we performed two sets of analyses. First, we preserved the time-based variability in the GPAs but constrained the variability in independent measures. In other words, we obtained the first principal components of all student- and teacher-generated indicators across time and ran a set of analyses similar to those described earlier: We let GPAs vary in time, but our predictors were all single-point indicators.

Second, we obtained an average GPA across the six terms and completed a set of traditional hierarchical linear regression analyses, evaluating increments in prediction over SSAT indicators and pre-Choate GPA.

Conditional Growth in Choate GPA: What Matters?

Conceptually, these analyses are similar to those performed in Study 1, but in this study, the PACE Battery contained more indicators.

The summary of these analyses is shown in Table 7. When single-predictor models were fitted, 13 of the 14 investigated indicators were found to be associated either with the intercept or with the slope of the Choate GPA growth model. When considered in a single model, however, a number of indicators appeared redundant. To lessen the redundancy of the information, the reduced model was fitted and the reduction of the number of the parameters did not worsen the model fit (χ^2 difference with 10 *df* was 13.95, *ns*). The final reduced model indicated that the Choate GPA intercept is predicted by pre-Choate GPA, SSAT Quantitative, and two PACE Battery indicators—Academic Self-Efficacy and School Life Inventory. The rate of growth of GPA was predicted by pre-Choate GPA, SSAT Quantitative, and Academic Motivation.

To summarize, these analyses attempted to differentiate growth profiles of Choate GPA and succeeded in doing so with both traditional (i.e., pre-Choate GPA and standardized tests) and novel (WICS- and SRL-based) measures. The remaining question, however, addresses the relative predictive validity of WICS- and SRL-based measures over and above the pre-Choate GPA and standardized tests.

Predicting Average (Freshman and Sophomore) Choate GPA

As we indicated earlier, the school provided preadmission data for this class of freshman. Specifically, SSAT indicators and pre-Choate GPAs were available for these comparative analyses. However, these indicators were available only for a single time point. Correspondingly, all other indicators were condensed from multiple points of measurement into summative indicators by means of principal-component analyses.[8]

We carried out two sets of analyses. In the first set, we applied a hierarchical regression analysis, in which the three SSAT indicators were entered first in a single step and then followed by (a) indicators of practical competence (Practical Reasoning and School Life Inventory), (b) indicators of creative competence (creative writing and scientific reasoning), (c) the indicator of ethical reasoning (as a partial measure of wisdom), and (d) the SRL-related indicators (Academic Self-Efficacy, Academic Locus of Control, and Academic Motivation). The second set was identical to the first set with one difference: SSAT indicators were entered simultaneously with pre-Choate GPA. The results of both sets of analyses are shown in Table 8.

The top portion of Table 8 shows that, overall, the introduction of WICS- and SRL-related indicators increased the amount of explained variance in the average Choate GPA by approximately 165%. Specifically, the amount of variance explained by all the indicators of SSAT was approximately14%, and the introduction of indicators only of practical competence almost doubled this amount (27.2%). Subsequently, both indicators of creative competence and ethical reasoning contributed to the explained variance, with self-report indicators bringing the values of R^2 to 37%. Much like the results of growth modeling, indicators that contributed the most to this profile were School Life Inventory and Academic Self-Efficacy.

The bottom portion of Table 8 repeats the first set of analyses but includes Pre-Choate GPA as a predictor at the first step. It appears that middle-school GPA is the best predictor across the board, explaining about29% of freshman and sophomore GPA. As is often the case, the best predictor of later GPA is earlier GPA. After SSAT and middle-school GPA, the WICS- and SRL-related measures together explain about 10% of the variance. In other words, they explain approximately as much as all three indicators of SSAT do when they are presented collectively but with no other predictors in the regression equation. When SSAT indicators are entered after middle-school GPA, they explain only an additional 8% of the variance, that is, less than what is explained by the PACE Battery measures with SSAT indicators and pre-Choate GPA in the regression equation.

In summary, the WICS- and SRL-related measures substantially increase predictive power for Choate GPA over SSAT alone, given, of course, that SSAT was used in the admission of students and thus was restricted in range.

Discussion

In this study, we generally confirmed and expanded the findings from our Pilot Study. Specifically, we showed, on a larger, more generalizable sample, that measures derived from the WICS and SRL theoretical frameworks substantially increased the predictability of Choate GPA. In addition, in conjunction with traditional measures of pre-Choate GPA and standardized tests, the measures of the PACE Battery predict not only the mean value of GPA but also its rate of growth. The magnitude of predictive validity is substantial, totaling up to 50% of the variance in Choate GPA. Although the best single predictor of Choate GPA appears to be middle-school GPA, measures from the PACE Battery predict as much as or more variance than do indicators of SSAT. They also contribute substantially, independent of SSAT.

[8] These are indicators of cumulative percentages of variance explained by the first principal component across multiple time measures: (a) Academic Self-Efficacy—63.7%; (b) Academic Locus of Control—71.4%; (c) Academic Motivation—75.2%; (d) School Life Inventory—74.4%; (e) practical reasoning task*:* quality of writing—63.7%, practicality—74.0%, quality of argument—56.2%; (f) creative writing—52.5%; (g) scientific reasoning—61.5%; (h) attribution of blame (58.7%).

Table 7
Conditional Change in Choate GPA Over Freshman and Sophomore Years at Choate (Study 1, Main Study)

Model	Intercept (starting value)			Slope (growth rate)		
	Coefficient	*T*	*p*	Coefficient	*T*	*p*
		Single predictor				
Parameter model						
Pre-Choate GPA	.9146	8.93	<.001	.0241	4.88	<.001
SSAT Verbal	.0078	2.77	.006	.0000	0.18	.855
SSAT Quantitative	.0130	4.71	<.001	.0006	5.07	<.001
SSAT Reading	.0070	2.23	.015	−.0001	−0.53	.595
Academic Self-Efficacy	.1374	3.49	.001	−.0012	−0.82	.415
Academic Locus of Control	.0704	1.74	.082	.0005	0.33	.744
Academic Motivation	.1532	3.93	.039	.0053	3.46	.001
School Life Inventory	−.1490	−3.79	<.001	−.0014	−0.88	.378
Practical reasoning						
Quality of writing	.0900	2.23	.026	.0015	0.97	.334
Practicality	.1308	3.31	.001	.0034	2.26	.024
Quality of argument	.0441	1.08	.280	−.0007	−0.48	.628
Creative writing	.1169	2.94	.004	.0011	0.69	.487
Scientific reasoning	.1673	4.32	<.001	.0015	1.01	.313
Attribution of blame	.1082	2.71	.007	−.0005	−0.34	.736
		Multiple predictor (combined)				
Full model						
Pre-Choate GPA	.6762	7.13	<.001	.0156	3.00	.003
SSAT Verbal	.0014	0.54	.587			
SSAT Quantitative	.0101	4.41	<.001	.0005	4.47	<.001
SSAT Reading	.0017	0.72	.471			
Academic Self-Efficacy	.0749	2.23	.025			
Academic Locus of Control	.0017	0.53	.958			
Academic Motivation	.0170	0.52	.605	.0028	1.69	.090
School Life Inventory	−.0780	−2.28	.022			
Practical reasoning						
Quality of writing	.0289	0.93	.351			
Practicality	.0303	0.93	.357	.0026	1.62	.105
Creative writing	.0170	0.52	.605			
Scientific reasoning	.0396	1.15	.251			
Attribution of blame	.0098	0.31	.759			
Reduced model						
Pre-Choate GPA	.7355	7.88	<.001	.0171	3.34	.001
SSAT Quantitative	.0105	4.84	<.001	.0005	4.22	<.001
Academic Self-Efficacy	.0978	3.28	.001			
Academic Motivation				.0036	2.31	.021
School Life Inventory	−.1151	−3.86	<.001			

Note. GPA = grade point average; SSAT = Secondary School Admission Test.

Study 2

On the basis of the results from Study 1, we made the decision to introduce selected measures from the PACE Battery into Choate's admission process. To minimize the associated costs and to avoid the need for proctoring, we included only self-assessments; they were presented to applicants as an "optional addition" to their application packages. In addition, during candidates' evaluations, the Choate admissions office representatives rated their perceptions of the creative and practical abilities of the interviewees, based on the interviewees' answers in the admission application to two short essay questions that were developed by Choate staff in concordance with the WICS theory to elicit a creative or practical response; the representatives used a scale ranging from 1 (*high*) to 3 (*low*), mimicking, at low cost, the performance measures used in Study 1. Like many measures in the admission process, this tool was used in an effort to either distinguish a candidate, positively or negatively, or to corroborate with other evidence in the application, whether it be the interview, other writing samples or the confidential teacher recommendations. All participation in this research was voluntary.

Method

The main objective of this study was to investigate the power to predict first-trimester Choate GPA on the basis of a number of conventional assessments (i.e., previous school GPA and SSAT) and augmenting assessments (i.e., self-report instruments from the PACE Battery).

Participants

For the academic year in which the study was conducted, Choate had 1,495 applicants (n = 685, or 45.8% female and n =

Table 8
Incremental Prediction of High School GPA Using WICS- and SRL-Related Measures (A) Above and Beyond SSAT and (B) Above and Beyond SSAT and Pre-Choate GPA (Study 1, Main Study)

Measure	Step 1	Step 2	Step 3	Step 4	Step 5
			A		
SSAT					
Verbal	.097	.052	−.009	−.011	.050
Quantitative	.304***	.374***	.398***	.397***	.386***
Reading	.071	.011	−.004	−.003	−.031
PACE WICS Indicators					
Practical competence					
Practical reasoning		.208**	.127	.125	.107
School Life Inventory		−.267***	−.209**	−.197*	−.191*
Creative competence					
Creative writing			.137	.134	.085
Scientific reasoning			.173*	.172*	.137
Ethical reasoning					
Attribution of blame				.031	.022
PACE SRL indicators					
Academic Self-Efficacy					.176*
Academic Locus of Control					−.012
Academic Motivation					.120
R^2	.142	.272	.318	.319	.374
			B		
SSAT					
Verbal	.057	.032	−.002	.004	.043
Quantitative	.193**	.250***	.270***	.269**	.271***
Reading	.136	.088	.075	.076	.048
Pre-Choate GPA	.552***	.494***	.468***	.467***	.440***
PACE WICS Indicators					
Practical competence					
Practical reasoning		.120	.078	.076	.072
School Life Inventory		−.201**	−.168**	−.158*	−.156*
Creative competence					
Creative writing			.072	.070	.038
Scientific reasoning			.109	.107	.083
Ethical reasoning					
Attribution of blame				.026	.019
PACE SRL indicators					
Academic Self-Efficacy					.147***
Academic Locus of Control					.010
Academic Motivation					.054
R^2	.433	.491	.506	.506	.535

Note. $n = 152$. Entries are standardized beta coefficients. GPA = grade point average; WICS = Wisdom, Intelligence, Creativity Synthesized; SRL = self-regulated learning; SSAT = Secondary School Admission Test; PACE = Psychology of Abilities, Competencies, and Expertise.
* $p < .05$. ** $p < .01$. *** $p < .001$.

810, or 54.2% male). Of these applicants, 377 (25.2% of the total sample, or 186 [or 49.3%] females and 191 [or 50.7%] of males) engaged the option of completing self-report assessments: the Academic Self-Efficacy Scale, the Academic Locus of Control Scale, and the Intrinsic/Extrinsic Academic Motivation Scale. The applicants who took the self-assessments did not differ from the rest of the sample on gender and age or SSAT Quantitative. Yet, those who took additional assessments tended to have slightly higher GPA at their previous school (3.67 vs. 3.55, $p < .001$), and slightly higher Reading (689.76 vs. 679.79, $p < .05$) and Verbal (705.13 vs. 694.43, $p < .01$) SSAT scores. Also, compared with the ethnic profile of the total pool of applicants, White students were overrepresented among those students who took the self-assessments (56% vs. 38.6% among those who took and did not take the assessment, respectively).

The school admitted 259 students (115 girls and 144 boys, mean age = 173.6 months, $SD = 16.06$),[9] 88 of whom took the self-assessments. These 88 students were predominantly White (61.7%), but many students (38.3%) came from a variety of other ethnic groups. Among the admitted students who took the assessments, girls were slightly overrepresented (53.4% vs. 39.8% among those who were admitted and did not take the assessment, respectively), but there were no ethnicity-based differences. In

[9] These students were admitted into a variety of classes and grades.

addition, although there were no SSAT-based differences among students with and without self-assessments, the GPA differences remained (3.92 vs. 3.62, $p < .001$).

Procedure

All self-reports were completed online. The applicant was requested to log in to a secure Web site, where the questions were presented electronically, the answers were recorded on a remote server, and the data were scored and delivered to Choate's admissions office. The admission decisions were made irrespective of the SRL indicators.

Measures

Self-reports. Applicants completed the Academic Self-Efficacy, Academic Locus of Control, and Intrinsic/Extrinsic Academic Motivation scales (see detailed descriptions in Study, Main Study). In this study, for the total sample ($N = 377$), Cronbach's αs were .80, .73, and .72 for the three scales, respectively. These alphas are comparable to those obtained in the Main Study of Study 1.

School reports. The school provided (a) demographic data (gender, age, ethnicity, and citizenship), (b) grades for all subjects the students took during their first trimester at Choate, summarized as fall GPA, and (c) preadmission data consisting of students' GPAs in their previous middle schools and their standardized test scores (i.e., SSAT Verbal, Quantitative, and Reading).

As mentioned earlier, admissions officers at Choate ranked each applicant on two dimensions, Creativity and Practicality. Each applicant was rated by two admissions officers. Their ratings were summarized by principal-component analyses, and the first component scores were saved for subsequent analyses. The first components explained 72% and 71% of the variance for Creativity and Practicality, respectively.

Results

Descriptive Statistics

Table 9 presents the descriptive statistics for the variables of interest (the PACE Battery indicators and the admissions officers' ratings, the SSAT indicators, and pre-Choate GPA) estimated for the whole sample of applicants and for the group of students admitted to Choate. First-trimester Choate GPA was, clearly, available only for the students ultimately admitted to Choate. As is apparent from Table 9, students admitted to Choate differed from nonadmitted students on pre-Choate GPA, $F(1, 1326) = 16.56$, $p < .001$, with the admitted students having higher GPA; SSAT Verbal, $F(1, 1293) = 10.00$, $p < .01$, with the admitted students having higher SSAT Verbal; and ratings of Creativity, $F(1, 1114) = 7.74$, $p < .01$, with the admitted students having lower ratings of creativity. There were no differences among admitted and nonadmitted students on the PACE SRL indicators; but again, because the completion of these self-reports was voluntary and the admission decisions were made irrespective of these indicators, the lack of group difference should be interpreted with caution, because possible biases cannot be ruled out.

When the demographic variables of gender and ethnicity status were considered, there were no group differences in the outcome variable (first-trimester Choate GPA). Thus, neither of these variables was included in subsequent analyses.

Predicting First-Trimester Choate GPA

Although the sample of volunteers who took self-report scales and were admitted to Choate was relatively small ($n = 88$), it provided us with an opportunity to investigate the predictive validity of pre-Choate GPA, the SSAT indicators, the PACE SRL indicators, and the admission ratings of Creativity and Practicality for first-trimester Choate GPA. Similar to the comparable regres-

Table 9
Means and Standard Deviations for Indices Used With Choate Applicants in Study 2

	Students who applied to Choate		Students admitted to Choate	
Sample	*M*	*SD*	*M*	*SD*
Indicator				
GPA				
Pre-Choate	3.58	0.48	3.69	0.41
Choate			3.28	0.42
SSAT				
Verbal	697.17	67.33	710.88	62.54
Quantitative	723.66	72.38	728.79	66.98
Reading	682.34	66.32	688.25	72.93
PACE SRL indicators				
Academic Self-Awareness	93.42	8.29	94.39	8.01
Academic Locus of Control	63.47	5.95	63.51	5.44
Academic Motivation	53.31	4.14	53.57	4.30
Admission rating				
Creativity	0.00	1.000	−0.17	0.98
Practicality	0.00	1.000	0.01	1.01

Note. GPA = grade point average; SSAT = Secondary School Admission Test; PACE = Psychology of Abilities, Competencies, and Expertise; SRL = self-regulated learning.

sion analyses conducted in Studies 1 and 2, we fitted two regression equations (see Table 10).

Three observations can be made on the basis of the results presented in Table 10. First, regression results here are similar to those from Study 1, presented in Table 3. Overall, the SSAT measures appear to explain only about 15% of Choate GPA, with SSAT Quantitative as the only variable making a statistically significant contribution to this prediction. This limitation may be in part a result of the restriction of range emanating from the use of SSAT scores in selection. The very brief self-reports forming the PACE SRL assessments contribute approximately 10% of unique predictive variance, and this contribution is sustainable with or without the presence of pre-Choate GPA. Second, in concert with earlier observations from both the Pilot Study and Main Study (Study 1), external Academic Locus of Control appears to be associated with higher Choate GPA. Third, in contrast to findings from the Main Study in Study 1, the indicators of Creativity and Practicality did not demonstrate predictive validity with regard to Choate GPA. This result underscores the importance of using stronger measures of these constructs, such as creative or practical maximum-performance tasks rather than self-reports and ratings by others, as has been done at the tertiary level (see Sternberg, 2007).

Table 10
Incremental Prediction of First-Trimester GPA Using SRL-Related Measures (A) Above and Beyond SSAT and (B) Above and Beyond SSAT and Pre-Choate GPA in Study 2

Measure	Step 1	Step 2
	A	
SSAT		
Verbal	−.190	−.175
Quantitative	.484**	.538**
Reading	.014	−.013
PACE SRL indicators		
Academic Self-Efficacy		−.008
Academic Locus of Control		−.330**
Academic Motivation		.100
Admission rating		
Creativity		−.048
Practicality		−.097
R^2	.150	.244
	B	
SSAT		
Verbal	−.186	−.167
Quantitative	.332*	.366*
Reading	.067	.042
Pre-Choate GPA	.419***	.419***
PACE SRL indicator		
Academic Self-Efficacy		.036
Academic Locus of Control		−.320**
Academic Motivation		.032
Admission rating		
Creativity		−.056
Practicality		−.114
R^2	.304	.402

Note. Entries are standardized beta coefficients. GPA = grade point average; SRL = self-regulated learning; SSAT = Secondary School Admission Test; PACE = Psychology of Abilities, Competencies, and Expertise.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

This study builds on the evidence obtained from both the Pilot and Main Studies (Study 1) by providing additional data suggesting that the predictive validity of middle-school GPA and standardized tests can be enhanced by the introduction of additional theory-based measures, such as self-reports, providing information on students' SRL. Although the gain in explained variance is relatively modest (i.e., about 10% of R^2), it is important to consider the costs, for both applicants and schools, in administering and basing their decisions on standardized tests, such as the SSAT, if three brief self-reports account for as much as approximately 67% of the variance accounted for by the SSAT (i.e., 10% vs. 15%). In short, although standardized tests provide important and valid information, it is crucial not to overstate their predictive power; after all, many other factors regulating students' learning at Choate, among them, self-efficacy and locus of control, contribute to the academic success of students in the highly competitive environment of college preparatory schools.

GENERAL DISCUSSION

The research described in this article shows that it is possible to construct an assessment, based on modern cognitive and motivational theories, that enhances the prediction of academic success over and above that of traditional tests and preentry GPA. The purpose of these assessments is not to replace but rather to supplement traditional tests. In SRL, learner attributes other than ability-achievement estimates are sought to augment the predictive power of prior grade and standardized tests. In WICS, analytical skills are theorized to be important for academic and life success, but other skills are assumed to matter as well. Thus, both analytical skills, whether assessed by the traditional measures for secondary school admissions or evaluated in some other reasonably objective way, and other skills need to be taken into account while predicting success in a secondary school.

An advantage of the PACE assessments is that they move educators and researchers beyond the drill-and-kill ritual that has come to pervade preparation for standardized tests. One could certainly develop skills that would enhance performance on our expanded assessments but, in doing so, would also be developing skills one needs for both school and life success.

Many independent schools are seeking new ways of improving their selection process, and even more importantly, schools are interested in changing the way they select to focus on the factors that matter most to the development of the person and the student process. Historically, it has been important to Choate to expand the traditional boarding school population. Rather than rely on "feeder" private schools for its student population, Choate has sought to expand the schools from which it accepts students. The broader and more diverse the sending school group, the more reflective of the general population and hence the broader the learning experience for the community. The research presented in this article might encourage other schools presently seeking to expand their cultural diversity to offer support similar to Choate's Icahn Program or to create other programs.

Another specific application of the findings presented in this article is toward counseling families with regard to their children's admission to Choate or a recommendation to look for a different

educational environment. When students flounder in the independent-school setting, it is upsetting for both the student and the faculty of the school; both feel disappointment and dissatisfaction. Our research suggests that it is possible to improve the prediction of academic success and, in some cases, avoid the wounds of student failure, at least in one school. We believe it may be worthwhile to try a similar approach in other schools. It might be possible to construct a battery that would work across many schools, or it might be better to customize batteries to individual schools. Whichever the case, the prediction of school success apparently can be enhanced by thinking more broadly about the skills that are measured at the time of application.

It is also important that, from a psychological point of view, our data show that modern psychological theories can improve the prediction of performance in an independent-school setting, beyond the prediction gained by a time-honored psychometric admissions test and entry GPA. Such customary measures provide some prediction. But our measures could potentially add to this prediction and show that some of the gap between the data desired and the data attained can be narrowed using measures based on modern psychological theories, such as the WICS (Sternberg, 2003), as well as modern theories of SRL (e.g., Boekaerts et al., 2000).

Although promising, these studies are characterized by a number of limitations. Specifically, the data presented here, although collected from many Choate students and applicants across a number of years, are still data collected from only one school. Thus, the degree of generalizability of these findings is difficult to judge before similar studies are carried out in other secondary schools. Another limitation of this work is that, although theoretically grounded, its theoretical scope is inclusive of only two modern theories of academic learning and success, SRL and WICS. Clearly, there are many more theoretical approaches in the literature that, when included in the conceptualization of the admission battery, may enhance its predictive power. Finally, though cohesive and consistent across all studies with regard to indicating the importance of predictors of academic success at Choate outside of GPA and conventional tests, the findings from Study 1 and Study 2 are slightly inconsistent. Although this inconsistency can be explained by the novelty of the measures, which could be strengthened, or the diversities of the samples including both enrolled and potential Choate students, clearly more research is needed to support and enhance the findings presented here.

Yet, we believe that this work is of interest, not only because it provides data supporting the importance of skills other than memory and analytical abilities for success in school but also because it expands the very limited research literature on admission and education in secondary schools. After virtually any study, there is always more to do, but this work, hopefully, formulates a convincing example for other secondary schools to experiment both with ways of carrying out their admission practices and diversifying their student populations.

References

Bandura, A. (1996). *Self-efficacy: The exercise of control.* New York: Freeman.

Beattie, S., Grise, P., & Algozzine, B. (1983). Effects of test modifications on the minimum competency performance of learning disabled students. *Learning Disability Quarterly, 6,* 75–77.

Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Eds.). (2000). *Handbook of self-regulation.* San Diego, CA: Academic Press.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning test* (No. 2000–1). New York: College Entrance Examination Board.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. New York: Psychology Press.

Facets. (2009). Facets (Version 3.65.0) [Computer software]. Beaverton, OR: Winsteps.

Hedlund, J., Wilt, J. M., Nebel, K. R., Ashford, S. J., & Sternberg, R. J. (2006). Assessing practical intelligence in business school admissions: A supplement to the graduate management admissions test. *Learning and Individual Differences, 16,* 101–127.

Legree, P. J. (1995). Evidence for an oblique social intelligence factor established with a Likert-based testing procedure. *Intelligence, 21,* 247–266.

Lupkowski-Shoplik, A. E., & Assoline, S. G. (1993). Identifying mathematically talented elementary students: Using the lower level of the SSAT. *Gifted Child Quarterly, 37,* 118–123.

Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development, 71,* 543–562.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3,* 97–105.

Mills, C. J., & Barnett, L. B. (1992). The use of the Secondary School Admission Test (SSAT) to identify academically talented elementary school students. *Gifted Child Quarterly, 36,* 155–159.

Múthen, L. K., & Múthen, B. O. (2005). *Mplus: Statistical analysis with latent variables. User's guide.* Los Angeles, CA: Author.

Noble, J., & Sawyer, R. (2002). *Predicting different levels of academic success in college using high school GPA and ACT composite score* (No. 2002–4). Iowa City, IA: ACT.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89,* 187–207.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2000). *HLM 5.* Lincolnwood, IL: Scientific Software International.

Rotter, J. B. (1990). Internal versus external control of reinforcement: A case history of a variable. *American Psychologist, 45,* 489–493.

Schuerger, J. M., & Dizney, H. F. (1967). The validity for ninth grade achievement of the SSAT and other admission criteria at a private secondary school. *Educational & Psychological Measurement, 27,* 433–438.

Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist, 40,* 85–94.

Sternberg, R. J. (1996). *Successful intelligence.* New York: Simon & Schuster.

Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized.* New York: Cambridge University Press.

Sternberg, R. J. (2007). Finding students who are wise, practical, and creative. *The Chronicle of Higher Education, 53,* B11.

Sternberg, R. J., & The Rainbow Project Collaborators. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence, 34,* 321–350.

Received January 23, 2008
Revision received March 2, 2009
Accepted March 11, 2009 ■