

Overview of College Board Noncognitive Work Carol Barry

Background

The College Board is well known for its work in successfully developing and validating cognitive measures to assess students' level of academic preparation; however, much less work has been devoted to measuring and studying noncognitive components of student success. By "noncognitive" we mean attitudes, beliefs, behaviors, and regulatory skills, all of which fall outside of the traditional domain of students' cognitive skills and abilities and which students need in order to be successful in college or career.

The literature indicates that to be successful in life, both cognitive and non-cognitive factors are important. Given the extensive body of research demonstrating the importance of non-cognitive constructs in student development, there is an opportunity for the College Board to conduct research and develop operational assessments in this domain. Measuring these noncognitive constructs would allow for a more complete understanding of students' strengths and weaknesses in academic settings. This information could subsequently be used to inform interventions designed to support students in these areas and to propel them toward college and career readiness. It is particularly important to understand strengths and weaknesses in these areas early in a student's academic career, while there is still time to improve weaknesses. For these reasons, The College Board has begun developing low-stakes assessments of noncognitive dimensions for middle, high, and postsecondary school students.

Our goal is to develop a **well-researched and well-developed combined assessment and intervention program** aimed at improving students' noncognitive skills. Specifically, this will identify students in need of intervention and support them to more successfully navigate the rigors of high school, complete high school ready for college, successfully transition to college, complete the first year of college, and return for a second year. This program is designed to serve **middle school, high school, and postsecondary students**.

We plan to develop **assessments that are innovative and interventions that are varied**. We envision combining quality noncognitive information with information from other College Board assessments (e.g., PSAT 8/9, PSAT/NMSQT, Accuplacer) to provide interventions/supports targeting areas of weakness. These interventions may be interwoven through instruction (e.g., incorporated in Pre-AP), may be delivered individually through individual assistance/supports (e.g., A20, postsecondary institutional supports), or delivered through Khan.

Overview of Work To Date

Establishment of Empirical Backbone

- Conducted a comprehensive literature review to determine which noncognitive dimensions are most strongly related to achievement, are particularly relevant for the intended grade level populations, and are malleable.
- Developed a Theoretical Framework for organizing and discussing these noncognitive dimensions and which will drive our instrument development. The framework groups relevant noncognitive dimensions into the following categories and describes how these categories relate to one another to impact achievement: Academic Behaviors, Academic Persistence, Academic Mindsets, and Learning Strategies.
- Established Empirical Backbone for middle school, high schools, and postsecondary assessments. This served as domain specifications for what was to be measured by the instrument, based on the criteria discussed above (i.e., most strongly related to achievement, relevance to middle school populations, and malleability).

Assessment Development

- Explored measurement methods and item types for noncognitive assessments. Specifically, reviewed noncognitive item types, appropriateness for population, appropriateness for constructs, amenability to computer-based testing, etc. We also worked with an external consultant and internal psychometrics staff to explore the use of forced-choice item types for our assessment.
- Development of Noncognitive Assessments and Content Validation
 - After adopting the theoretical framework, items were written to measure each of the included constructs. To this end, existing items from an earlier noncognitive assessment (i.e., the College Board's Study Skills Inventory) were reviewed/revised and additional new items were written to ensure adequate coverage of construct domain. Pilot forms of noncognitive assessments, including background items, noncognitive scales, complete test instructions, etc. were created.
 - After items were written, a content validation exercise was conducted for two purposes: 1) to gather evidence on whether items measured the intended construct or other constructs; and 2) to identify items that needed additional revision or clarification. Two independent raters, who had not participated in any prior item development efforts but were well-versed in the research and measurement of noncognitive constructs, mapped each item to the construct(s). Any discrepancies in mappings were discussed, some items were revised, and final item-mappings were reached.
 - Activities were first undertaken to develop a middle school assessment, but have since been conducted to develop high school and college versions of the instrument.

Pilot Testing

- Middle School Field Trial 1
 - Approximately 2500 8th graders from 26 middle schools who also participated in the redesigned PSAT 8/9 scaling study field trial
 - Schools administered the online noncognitive assessment within four weeks of the PSAT 8/9 administration
 - Data collection occurred between December 2014 and March 2015
 - Full data analyses were conducted including:
 - Data cleaning and screening
 - Examination of internal structure
 - Scale descriptive statistics and reliability estimates
 - Invariance/DIF analyses
 - Examination of relationships among noncognitive dimensions and with achievement
 - IRT analysis of data to aid in creation of forced-choice version of assessment
- Middle School Cognitive Interview Study 1
 - Conducted think aloud interviews (TAIs) for items with about 15 students
 - Used think aloud data to supplement the quantitative analysis of pilot data.
 - The TAIs provided insight into items that functioned poorly, which was then used to further refine the items and scales.
 - TAI data also coded for correct/incorrect interpretation to inform whether students correctly interpret/respond to items.
 - Collected desirability ratings for items with about 30 students
 - These desirability ratings were used to create forced-choice pairs.
- Middle School Field Trial 2, forced-choice assessment
 - Approximately 5,500 students from 18 middle schools completed the assessment.
 - Schools administered the online noncognitive assessment within four weeks of their PSAT 8/9 administration
 - Data analyses include:
 - Exploration of IRT methods to score forced-choice data
 - Evaluation of reliability
 - Examination of relationships between noncognitive scores and achievement scores
- High School Field Trial 1
 - Approximately 3200 10th graders from 7 high schools completed the complete noncognitive assessment. These students also took the PSAT 10.
 - Schools administered the online noncognitive assessment within four weeks of the PSAT 10 administration

- Full data analyses included:
 - Data cleaning and screening
 - Examination of internal structure
 - Scale descriptive statistics and reliability estimates
 - Invariance/DIF analyses
 - Examination of relationships among noncognitive dimensions and achievement
 - IRT analysis of data to aid in creation of forced-choice version of assessment (for later pilot)
- Middle School Cognitive Interview Study 2
 - Conducted think aloud interviews (TAIs) for items with about 15 students
 - Currently using think aloud data to supplement the quantitative analysis of pilot data.
 - TAI data are also being coded for correct/incorrect interpretation to inform whether students correctly interpret and respond to items.
- High School Cognitive Interview Study 1
 - Conducted think aloud interviews (TAIs) for items with about 15 students
 - Using think aloud data to supplement the quantitative analysis of pilot data.
 - The TAIs provided insight into items that functioned poorly, which was then used to further refine the items and scales.
 - TAI data are also being coded for correct/incorrect interpretation to inform whether students correctly interpret and respond to items.
 - Collected desirability ratings for items with about 30 students
 - These desirability ratings will be used to create forced-choice item pairs.

Examination of Student Profiles

- Used latent profile analysis to explore the existence of profiles to determine whether there are specific groups of students that have certain patterns of noncognitive and cognitive skills.
 - Preliminary analysis was of middle school pilot 1 data (i.e., Likert), but we have plans to analyze both middle school pilot 2 data (i.e., FC) and high school pilot 1 data (i.e., Likert)
- Our preliminary analysis provides evidence of several common types of students with different profiles of noncognitive and cognitive attributes.

Intervention Development

- Have begun development work for interventions designed to bolster students' noncognitive skills.
- Current have four interventions drafted:
 - Growth mindsets video
 - Values affirmation
 - Post-exam error analysis
 - Intentional change project
- Planning to pilot 1-2 of these interventions at small scale to determine impact on students' noncognitive levels.

Current and Future Work

- Continued exploration of forced-choice assessment as a feasible way to mitigate socially desirable responding.
- Continued exploration of assumptions underlying noncognitive assessment.
- Continued assessment development and pilot activities
 - HS pilot 2
 - Postsecondary Pilots
- Continued intervention development and pilot activities
 - Initial pilot of several interventions planned
- Long term validation research agenda

Theoretical Framework (Noncognitive Empirical Backbone)

School and Classroom Context

