Contemporary Educational Psychology 34 (2009) 195–209

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Using the theory of successful intelligence as a framework for developing assessments in AP physics

Steven E. Stemler [a,*], Robert J. Sternberg [b], Elena L. Grigorenko [c], Linda Jarvin [b], Kirsten Sharpes [a]

[a] Wesleyan University, 207 High Street, Middletown, CT 06549, United States
[b] Tufts University, Ballou Hall, Medford, MA 02155, United States
[c] Yale University, Child Study Center, 230 South Frontage Road, New Haven, CT 06519, United States

ARTICLE INFO

Article history:
Available online 5 May 2009

Keywords:
Intelligence
Achievement
Advanced placement
Physics
Cognitive processing

ABSTRACT

A new test of Advanced Placement Physics, explicitly designed to balance both content and cognitive-processing skills, was developed using Sternberg's theory of successful intelligence. The test was administered to 281 AP Physics students from 10 schools during the 2006–2007 school year. Six empirically distinguishable profiles of strengths and weaknesses emerged from an exploratory Q-type factor analysis across the four cognitive-skill areas assessed (i.e., memory, analytical, creative, and practical skills). These profiles replicated those found in previous research in the domains of AP Psychology and AP Statistics. Furthermore, achievement differences between ethnic groups on various cognitive subscales were reduced as compared with traditional estimates. The results provide evidence of the importance of integrating modern theories of cognitive processing into large-scale assessments.

1. Introduction

Each year, millions of students across the country take high-stakes achievement tests that will have an important influence on their academic and professional futures (Heubert & Hauser, 1999); yet, many of these tests are not aligned with modern theories of student learning and cognitive processing. As a result, students with strengths in cognitive skills not assessed by these tests may have their future opportunities curtailed (Sternberg, 1997). Indeed, many tests that serve as gatekeepers tend to emphasize only a limited range of skills (e.g., analytical and memory skills). Yet analytical and memory skills alone are not sufficient to succeed in the professional world. For example, although analytical skills are important to the physicist, who must compare and contrast competing explanations for phenomena and critically analyze data, other skills are important as well. It takes creative skills for the physicist to synthesize disparate findings and generate new theories, and practical skills to understand how theoretical findings may be used in the real world (e.g., to improve communication technology) as well as to persuade others of the value of the findings. To the extent that selection tests are weighted more heavily in favor of one particular type of skill, an entire professional field may suffer because it potentially will be dominated by individuals with a single profile of strengths and weaknesses, thereby inhibiting the capacity of the field to develop to its full potential. A balance of cognitive skills is important, regardless of one's professional domain. Thus, measurements should assess a broad profile of skills in students.

The aim of the current research was to examine the impact on student achievement of creating a set of modified, theory-driven examinations that expanded the range of cognitive skills assessed. The College Board's Advanced Placement (AP) program in Physics was used as a testing ground for the project.

1.1. The Advanced Placement Program

The College Board's Advanced Placement (AP) Program, initiated in 1955, was originally designed as a mechanism for granting exceptional high school students an opportunity for advanced study that would be equivalent to college-level programming. When this program began, it served only top students from a limited number of high schools, but in 2006, 666,067 graduating seniors (24% of all graduating seniors) at 16,000 secondary schools reported having taken at least one exam in one of the 37 courses across 22 subject areas offered by the AP program (College Board, 2007).[1]

* Corresponding author.
E-mail address: steven.stemler@wesleyan.edu (S.E. Stemler).

[1] The courses offered by the AP Program are: Art History, Biology, Calculus AB, Calculus BC, Chemistry, Chinese Language and Culture, Computer Science A, Computer Science AB, Macroeconomics, Microeconomics, English Language, English Literature, Environmental Science, European History, French Language, French Literature, German Language, Comparative Government & Politics, US Government & Politics, Human Geography, Italian Language and Culture, Japanese Language and Culture, Latin Literature, Latin: Vergil, Music Theory, Physics B, Physics C, Psychology, Spanish Language, Spanish Literature, Statistics, Studio Art: 2-D Design, Studio Art: 3-D Design, Studio Art: Drawing, US History, and World History.

0361-476X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cedpsych.2009.04.001

Each spring, students enrolled in AP courses are given the opportunity to take a high-stakes examination to demonstrate their mastery of the subject area. The exams are graded on a scale from 1 to 5, with a score of five indicating a student who is extremely well-qualified to receive college credit and/or advanced placement based on an AP exam grade (College Board, 2004). Most colleges will grant credit to students scoring three or higher on the exam. Thus, the results of the test have important financial implications, as placing out of an introductory college courses could potentially save a student thousands of dollars in tuition in subsequent years. In addition, AP scores are frequently used in admissions decisions as predictors of college success (Morgan & Ramist, 1998). The limited number of chances to take the test, the potentially significant financial savings associated with the outcome, and the impact scores may have on college admissions decisions qualifies the AP examination as a high-stakes test that has a broad impact on hundreds of thousands of high school students each year.

Historically, the chief concern of AP exam developers has been with ensuring adequate content-area coverage. For example, the items on the AP Physics B exam are explicitly balanced to ensure proportionate representation of various subtopics within the domain of Physics (i.e., Newtonian mechanics; fluid mechanics and thermal physics; electricity and magnetism; waves and optics; and nuclear physics). Traditionally, however, there has been no systematic attempt explicitly to balance items for the cognitive-skill areas they assess.

1.1.1. Ethnic differences in achievement

One of the biggest challenges facing the AP program is in the recruitment of minority students to participate in the program. Research has found that African–American and Latino students enroll in AP courses at approximately half the rate of White students. In particular, minority students enroll in AP math, science, and English classes at lower rates than White students at comparable schools (Klopfenstein, 2004; Ramist, Lewis, & McCamley-Jenkins, 1994). As a result of this differential enrollment, fewer minority students end up taking AP exams. In 2006, approximately 21% of all students who took one or more exams were African–American or Latino; by way of comparison, approximately 30% of students enrolled in high schools were African–American or Latino (College Board, 2007). Because taking an AP course is a strong predictor of whether a student will take an upper level class or major in that subject in college (Dodd, Fitzpatrick, DeAyala, & Jennings, 2002; Morgan & Maneckshana, 2000), the AP courses that students choose to take have important implications for their future course of study and, eventually, their profession.

In addition to the problem of low minority-student enrollment in advanced courses, one of the most persistent problems in instruction and assessment over the years has been the existence of systematic differences in student achievement across ethnic groups (Chubb & Loveless, 2002; Jencks & Phillips, 1998). Indeed, research suggests that White students receive higher scores on standardized tests than African–American, Latino, and Native-American students as early as preschool (Nettles & Nettles, 1999). This difference is dramatic on most conventional achievement tests; nearly a full standard deviation separates the average scores of African–American and White high school students (Hedges & Nowell, 1998). This pattern holds for scores on the AP exam as well. For example, in 2006 the mean score for African–American test-takers across all AP exams was 1.96, compared with 2.96 for White students (College Board, 2007). This difference is not only large but consequential: because three is typically a passing score for getting college credit, the average White student will "pass" a given AP exam while the average African–American student will "fail" it.

The difference in scores of students from different ethnic backgrounds is more dramatic in some domains than in others. For example, there is little difference between the scores of White students and African–American students on the AP Studio Art: 3D-Design exam; the average score of African–American students was 2.68 compared with 2.95 for White students. But a difference of 1.13 separates the average scores of White students and African–American students on the AP Physics C exam; comparable results are 1.15 for AP Microeconomics, and 1.35 for AP Computer Science scores (College Board, 2007). As AP scores are a useful indicator of college success and an important consideration in the college-admissions process, differences in these scores have high-stakes consequences.

Researchers have proposed several possible reasons for the achievement gap between White students and underrepresented minorities, including genetic differences (Herrnstein & Murray, 1994), cultural differences (Fordham & Ogbu, 1986; Williams, 2004), social–psychological differences (Steele, 1997), and differences in the quality of instruction (Hanushek & Rivkin, 2006). Another potential reason for this persistent difference, however, is that traditional achievement tests have assessed a fairly limited range of cognitive processes, ignoring other important skills.

Sternberg and colleagues have demonstrated in a series of studies that when assessments are designed in such a way that they expand the range of cognitive skills assessed, the achievement gap between White students and minority students is reduced. For example, in a recent study designed to create assessments that would enhance the predictive power of the SAT, Sternberg and The Rainbow Project Collaborators, (2006) found that adding assessments of creative and practical skills doubled the power of the battery to predict first-year college GPA compared with the use of the SAT alone. In addition, differences in achievement between White and African–American students were reduced on measures of creative skills, and differences in achievement between White and Latino students were reduced on assessments that emphasized practical skills and creative skills.

The decrease in the achievement gap as a result of measuring a broader range of cognitive skills has also been demonstrated in the context of the AP program. Stemler, Grigorenko, Jarvin, and Sternberg (2006) designed augmented versions of the AP Psychology and AP Statistics examinations that included practical and creative subscales. A key finding was that the effect-size difference between African–American students and White students was virtually nonexistent for both the creative subscale ($d = -0.02$) and the memory subscale ($d = 0.04$) of the modified exams. The largest difference between Latino students and White students was observed on the memory subscale of the modified AP Psychology exam, in which Latino students scored approximately one-half a standard deviation below the White students ($d = -0.47$). Yet, the effect-size difference between Latino students and White students was somewhat lower on the creative subscale ($d = -0.32$), and substantially lower on the practical subscale ($d = -0.13$). Results of the AP Statistics exam showed a similar pattern. Overall, the findings from these past studies suggest that developing assessments that measure a broad range of cognitive abilities may help to create more equitable achievement tests.

1.2. Theoretical framework

In recent years, designers of large-scale testing programs, recognizing the important social, economic, and ethical consequences associated with standardized testing, have become particularly interested in linking educational assessment to modern theories of cognitive-processing skills (Embretson & Reise, 2000; Irvine & Kyllonen, 2002). Capitalizing on this idea, the current project involved the development of an augmented test in the subject area of AP Physics B that was explicitly linked to Sternberg's (1997,