

POLICY NOTES

News from the ETS Policy Information Center

Volume 17, Number 1

Policy Evaluation & Research Center

Winter 2009

Addressing Achievement Gaps

Educational Testing in America: State Assessments, Achievement Gaps, National Policy and Innovations



U.S. Secretary of Education Margaret Spellings provided the keynote address

in America: State Assessments, Achievement Gaps, National Policy and Innovations,” a recent conference cosponsored by ETS and the College Board. “The achievement gap starts at birth and follows students all the way through high school, and we have a moral responsibility to do something about that.”

But whether the NCLB-mandated assessment system, under which states test schoolchildren annually in reading and math and report the results by demographic subgroups, has helped or hurt the effort to close achievement gaps between rich and poor, minority and White, remains a complicated and difficult question, argued speakers at the conference, the 11th in ETS’s series of “Addressing Achievement Gaps” symposia.

NCLB’s greatest contribution, speakers said, is the spotlight it has turned on the achievement of demographic subgroups, whose under-performance used to lie hidden within school district and state averages. That new attention has brought extra help to struggling students and long overdue attention to the national challenge of ensuring equal educational opportunity to students of all backgrounds.

Seven years after the federal No Child Left Behind Act (NCLB) put testing at the top of the nation’s education agenda, policymakers and reformers on both the right and the left agree that achievement gaps based on race, ethnicity and class must close if the United States is to maintain its economic pre-eminence and live up to its founding principles. “We are creating a larger and larger cohort of socioeconomic demographically disadvantaged children in this country,” ETS’s President and CEO **Kurt M. Landgraf** said as he opened the symposium “Educational Testing

THIS ISSUE

Educational Testing in America: State Assessments, Achievement Gaps, National Policy and Innovations

Annual standardized testing lies at the heart of the accountability system that American education reformers and policymakers have established during the past decade in an effort to ensure equal opportunity for all students, no matter their race, ethnicity or wealth. The new testing regime has brought national attention to the schooling of disadvantaged children, and in some states and school districts, achievement gaps between low-income and minority students and their middle-class, White peers have begun to narrow. Critics charge, however, that high-profile annual testing has also shaped the education system in ways that sometimes hurt the very students who most need help. And educators and policymakers have begun to realize that the essential task of closing achievement gaps will require new kinds of accountability systems, and new kinds of tests.

(continued on back page)

“We have looked ourselves in the mirror and have focused as never before on the achievement of every child,” U.S. Secretary of Education Margaret Spellings said in her keynote address. “And that’s right, and it’s righteous.”

‘We have looked ourselves in the mirror and have focused as never before on the achievement of every child. And that’s right, and it’s righteous.’

— *Margaret Spellings*

But the new emphasis on accountability through testing has also had undesirable side effects, speakers said. States trying to put the best political face on test results have set proficiency-score cutoffs so low that even students who pass need remedial help before they can do college work. The focus on reading and math has narrowed the curriculum in some schools, depriving disadvantaged children of enriching academic experiences. And although only effective teaching will narrow score gaps, annual tests paradoxically give teachers little help in tailoring instruction for failing students.

“Assessments can help close achievement gaps,” **Brian Gong**, the Executive Director of the National Center for the Improvement of Educational Assessment, told symposium participants. “But they can’t do it by themselves, and they can’t do it within their current structure.”

‘Assessments can help close achievement gaps. But they can’t do it by themselves, and they can’t do it within their current structure.’ — *Brian Gong*

The solution is not to jettison annual standardized tests or the proficiency demands they embody, symposium speakers said. Rather, speakers argued, the solution is to broaden our accountability system beyond once-a-year tests of cognitive skills — by refining the curriculum standards on which testing is based, by developing tests that help teachers improve the instruction they provide, and

by finding new ways to assess the noncognitive skills that students need for success in college and the workplace.

Many States, Many Tests

State testing has changed dramatically in the past decade, researcher **Laurens L. Wise** of the Human Resources Research Organization (HumRRO) told the audience in the symposium’s opening session. Since the passage of NCLB, the number of state tests has exploded, and these tests are increasingly used to make high-stakes decisions about grade promotion and high school graduation. The new landscape has many positive features, Wise argued. Policymakers, not test developers, now decide what students should learn; test validity information is closely scrutinized; and test results are reported and discussed widely. Pegging scores to academic standards, rather than to the performance of other test takers, allows meaningful discussion of whether achievement levels are high enough, he said. And NCLB’s requirement that states report test results by demographic subgroups has brought new attention to the achievement gap. With this increased attention has come increased help for struggling students, Wise said, and a narrowing of the White-Black and White-Hispanic score gaps on the National Assessment of Educational Progress (NAEP).

But the state assessment regime has significant shortcomings, Wise said. Although academic standards laying out what skills and knowledge students must acquire in 13 years of schooling are the foundation of the accountability system, states seldom explain why their standards include particular content or collect data to support those determinations. Even states that align content standards vertically, from grade to grade, do not rely on empirical evidence to explain why mastering the required material at one grade level

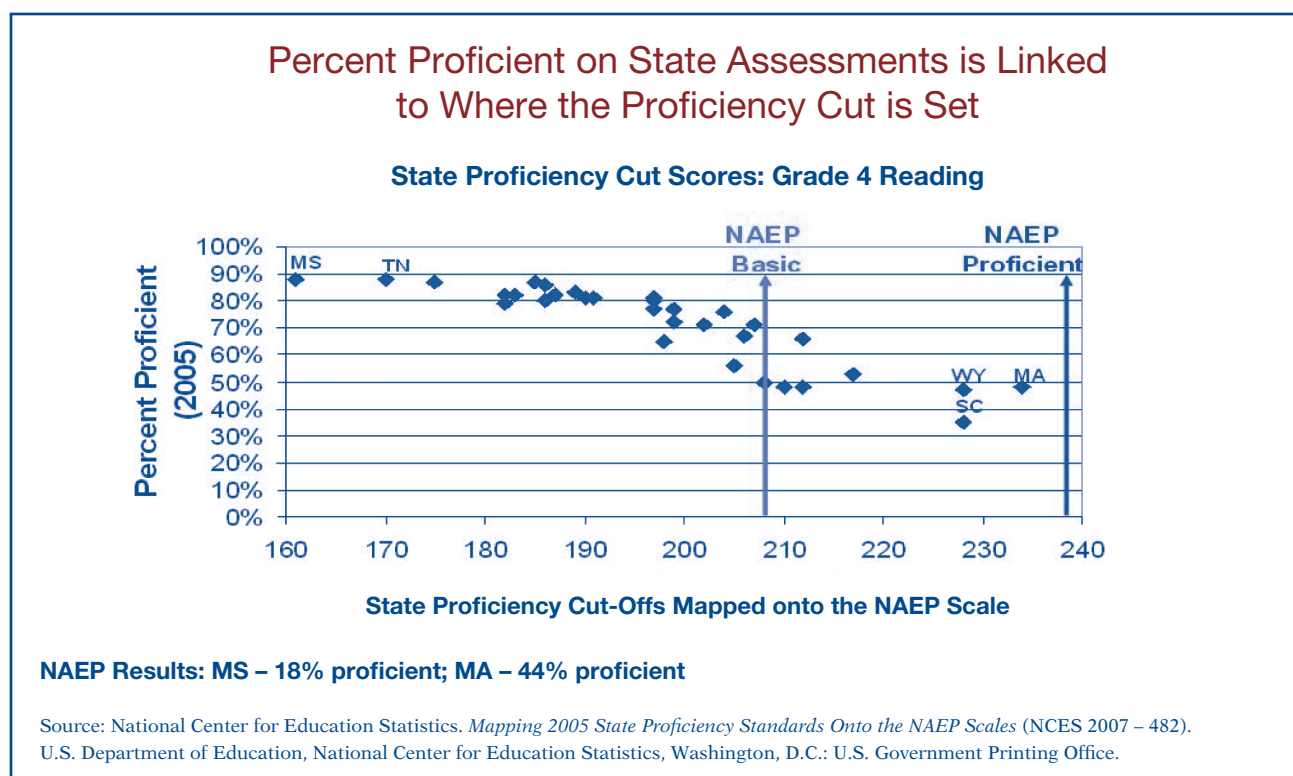
is a prerequisite for understanding next year’s work. And states seldom consider what other states — let alone other countries — expect from their students. Lacking a data-driven rationale for their content standards, Wise said, states tend “to just throw everything in,” making it difficult to design tests that fully assess all the required content.

Problems with standards are matched by problems with tests, Wise said. The proliferation of tests, each customized to fit a different set of state standards, spreads developers thin, and the money spent to give each state its own test of, say, fifth-grade math might be better spent on math instruction. In years past, when Kansas children grew up to raise corn and Pittsburgh children grew up to forge steel, giving localities wide latitude in the skills and knowledge they demanded from students made sense, Wise said, but in an era of geographic mobility and international competition, “it’s not clear that makes as much sense today as it once did.”

‘Is it fair to the students in Mississippi to expect so much less of them than we expect of the students in Massachusetts? Who’s looking at the between-state achievement gaps?’

— *Laurens L. Wise*

Not only have states adopted different tests; they have also defined proficiency on those tests in vastly different ways, sometimes sticking close to the proficiency standard required by the widely respected NAEP, but sometimes setting a far lower bar in order to produce a more politically palatable success rate (see the graph below). Those differences raise equity questions, Wise said. Ninety percent of Mississippi’s students are deemed proficient on the state’s test, but only 18 percent meet the NAEP standard; meanwhile, in Massachusetts, 50 percent of students meet the state’s proficiency threshold, a closer fit with the state’s 44 percent NAEP proficiency rate. “Is it fair to the students in Mississippi to expect so much less of them than we expect of the students in Massachusetts?” Wise asked. “Who’s looking at the between-state achievement gaps?”



The world after high school offers further evidence that proficiency-score cutoffs are political compromises, rather than meaningful measures of achievement, speakers argued. Even students who achieve proficiency on state tests often need remedial instruction before they can do college work, and, as a result, colleges spend \$1.4 billion a year providing that remediation, said **Youlonda Copeland-Morgan**, a Syracuse University administrator and the Board of Trustees Chair-elect at the College Board. “We’re talking about pretty modest levels of performance that are in no way a representation of what proficiency means by our conventional definitions,” said ETS researcher **Drew H. Gitomer**. Whatever the definition of proficiency, NCLB’s standards should not be the sole measure of educational effectiveness, said **David P. Cleary**, a staff member for Republican U.S. Senator Lamar Alexander of Tennessee. Meeting NCLB requirements signifies only that a school system does not need federal intervention, Cleary said: “You can have really good scores and still not be a great school.”

The shortcomings in the current testing regime have implications for efforts to close the achievement gap, speakers pointed out. If state standards bear only an imperfect relation to real-world demands, tests measuring mastery of those standards may not highlight the achievement gaps that really need closing; if proficiency cutoffs are set artificially low, getting every student over that low bar will not ensure workplace success and international competitiveness. The challenge, said **Mitchell D. Chester**, the Massachusetts Commissioner of Elementary and Secondary Education, is “anchoring our notions of what’s good enough, our performance standards and our content standards, in some real-world criteria.”

A Closed Loop

If the current accountability system faces problems at the policy level, it has also spawned unintended consequences inside classrooms. NCLB’s focus on reading and math scores has convinced some schools, especially those serving the low-income and minority students who struggle hardest to reach proficiency, to narrow their curricula to a drill-based march through the three Rs, eliminating subjects such as art, music and physical education, speakers said. “Too often, the state test is turned to as the curriculum,” said **Roberto Rodriguez**, a staff member for Democratic U.S. Senator Edward M. Kennedy of Massachusetts. Indeed, defining success by reference to a single proficiency score encourages an even more radical curricular narrowing, said **John Tanner**, Director of the Center for Innovative Measures at the Council of Chief State School Officers. To achieve adequate proficiency scores, schools need never teach the simplest material (since students will get the easy questions right anyway) or the most complicated (since students who get the hard questions wrong will still pass the test). Instead, Tanner said, struggling schools may choose to teach only the mid-level content, in hopes of boosting as many students as possible over the all-important proficiency line.

‘Too often, the state test is turned to as the curriculum.’ — Roberto Rodriguez

Despite reformers’ best intentions, using test scores as the gauge of school success has distorted the educational system, Tanner said. Standardized test scores were supposed to serve as proxies for something outside the test — literacy, numeracy, workplace skills — but the proxy has become an end in itself. “Standards and assessments now function as a closed loop,” Tanner said. “We ask

if we were successful within the closed loop, but we also know that there are so many other things critical to success.”

‘Standards and assessments now function as a closed loop. We ask if we were successful within the closed loop, but we also know that there are so many other things critical to success.’

— *John Tanner*

Is this narrowing of schools’ horizons an inevitable result of NCLB’s accountability regime? Not surprisingly, Secretary of Education Spellings disputed that notion. “It’s the expectation for our own children that they read and cipher on grade level and, oh, yeah, they have P.E. and art, too,” she said. “Why are these things mutually exclusive?”

Other speakers, however, portrayed a narrowed curriculum as a logical result of the accountability that the NCLB testing regime demands from teachers and schools: “We are getting exactly what we designed the system to do, inadvertently,” Tanner said. The challenge, speakers agreed, is to create a new system that retains reformers’ strong commitment to closing achievement gaps but that avoids the pitfalls of the current regime. Connecticut, for example, spurred schools to offer a richer science curriculum by administering a 10th-grade science test that included questions about a classroom lab experiment students had to perform six weeks earlier, said Massachusetts Commissioner Chester. “The inference that folks are reaching on the ground in too many cases is that the way to prepare for the test is to drop what you would think of as a regular curriculum and come up with this narrow, more focused, test-preparation type of scheme,” Chester said. “How can we design state assessment systems that create some evidence for teachers that if their day-to-day curriculum is much more aspirational, they will in fact be preparing kids for the tests?”

‘How can we design state assessment systems that create some evidence for teachers that if their day-to-day curriculum is much more aspirational, they will in fact be preparing kids for the tests?’

— *Mitchell D. Chester*

An accountability system based on a single year-end test has another shortcoming, speakers said: such tests give teachers little guidance in the day-to-day work of helping struggling students master state standards. Surveying years of state and national test score data, Gong concluded, “We could spend a lot of time looking at that, and we still don’t get very much information about what informs our action, particularly at the district, school or classroom level.” And the classroom is the only place where achievement gaps can be not merely identified, but closed, said **Rick Stiggins**, the Executive Director of ETS’s Assessment Training Institute. “The bottom line is that only teachers can use assessment day to day to support the learning of their students,” Stiggins said. All too often, however, neither teachers nor principals are trained to use assessment effectively, he said. Other speakers echoed the point. In Maine, said state Commissioner of Education **Susan A. Gendron**, legislators repealed a law incorporating locally designed assessments into the state accountability system because teachers lacked the “assessment literacy tools” to make the plan workable.

‘The bottom line is that only teachers can use assessment day to day to support the learning of their students.’ — *Rick Stiggins*

If teachers do not get what they need from our current testing system, most students get even less, Stiggins said. Although the intimidating ordeal of an annual pass-fail proficiency assessment may motivate some students, it leaves others discouraged and hopeless. “If all students

are to meet standards, then they must all believe they can, because if they don't believe that, there isn't going to be any achievement-score gap-closing," Stiggins said. "You don't fix that with another \$100 million statewide testing program. You fix this in the classroom."

Balanced Assessment Systems

The solution to the problems of the current testing regime is not an end to that regime, and still less to its call for holding all students to the same standards, symposium speakers stressed. "We don't want to replicate the system of the past," Massachusetts official Chester said. "The system of the past was, what was good enough in District A would never qualify as good enough in District B. And that cheated kids in District A." Instead, speakers said, we need to refine our academic standards, redesign our assessment regime to answer a larger set of questions, and develop new kinds of tests that assess new kinds of skills.

Improving content standards is essential to the enterprise, Gong said. Currently, state standards often do not spell out every element of what students need to know to achieve proficiency, he said. A math standard, for instance, may call for students to partition an area into parts and then identify the fraction described by the partitioned area, but teachers will need to ensure that students have mastered a number of basic concepts — such as the difference between part and whole — before even beginning the exercise; standards should include detailed learning progressions spelling out these prerequisites. States also need to lay out the steps by which students progress along the path toward mastering standards, Stiggins said, since mastery is a gradual process of development. "How do you close the achievement gap without a vision of the continuum along which the gap exists?" he asked.

Any assessment system that aims to close achievement gaps must also include more than a single year-end test, no matter how well designed, speakers said. An assessment system must answer many questions, Stiggins said: policymakers need to know how many students are meeting standards, in order to hold schools accountable; district officials need to know which standards their students cannot meet, in order to design better programs; and teachers need to know what material their students have not yet mastered, in order to decide what to work on next. The current state testing system answers only the policymakers' questions, Stiggins said, but "in a balanced accountability system, we conduct assessments in a manner that answers all of the critical questions, not just some of them." Thus, a balanced assessment system would include not only annual standardized tests providing political accountability, but also periodic benchmark assessments designed to gauge the success of programs and frequent classroom tests aimed at diagnosing the problems of individual students.

Educators are beginning to respond to these new imperatives, according to Gong and Stiggins. Districts have created uniform pacing guides that tell teachers how quickly to cover material, and some school systems administer interim assessments to measure how well students are learning the material the state test covers. But these new tests are problematic, Gong said, since few have been reviewed for quality and many simply mirror the content of the corresponding year-end test. Interim assessments covering material that teachers have not yet taught provide little useful diagnostic information, he noted. To help teachers improve their practice, Gong said, interim assessments must gauge student progress relative to the detailed learning progressions contained in refined state standards.

Districts must also pay attention to students' course-taking patterns, speakers noted. In one Delaware high school, most low-income students took only low-level math classes. "Now I think I know why they've got the results that they do in terms of the state math test," Gong said. Students with disabilities and English-language learners also often miss out on crucial coursework, HumRRO researcher Wise said. "Not surprisingly," he said, "if they're not being instructed in the materials covered by the test, they don't pass."

New Kinds of Measures

In a new, improved assessment regime, tests would not only document students' learning and help teachers improve their instruction, but the tests themselves would also offer worthwhile educational experiences, said Gitomer, Senior Director of ETS's Center for the Study of Teacher Assessment. In middle schools in Portland, Maine, ETS is developing such assessments — known as Cognitively Based Assessments *of, for, and as* Learning, or CBAL — in reading, writing and math. Unlike traditional standardized tests, CBAL builds on cognitive-science research about how learners achieve proficiency. Standard comprehension tests, for example, assess only some of the skills required for reading proficiency, Gitomer said, ignoring both the basic prerequisites of comprehension, such as the ability to decode text, and the more sophisticated interpretative methods that proficient readers apply to different kinds of texts. CBAL tries instead to test the full range of required reading skills and to embed that assessment in educationally meaningful tasks.

CBAL vs. (stereo)typical assessments

Traditional	CBAL
Single measurement occasion	Multiple measurement occasions
Many short items (mostly) unrelated Representative of a domain	A few long tasks Centered around a common theme Based on a competency model
Homogeneous response types	Heterogeneous response types

Source: Educational Testing Service.

To accomplish these broader goals, Gitomer explained, CBAL replaces the traditional days-long testing marathon with a series of shorter tests — Periodic Accountability Assessments (PAAs) — that are given throughout the school year and thus can provide teachers with useful information about students' progress.

A reading PAA could begin with a spoken module requiring the test taker to read aloud into a headset, with a computer scoring for accuracy and fluency, basic prerequisites of reading comprehension. Because middle schools often assume students have mastered these basics, a teacher using a traditional reading comprehension test might conclude that a failing student needed more help with comprehension; by contrast, the PAA can detect students who are struggling at an even more basic level.

The reading PAA might continue with a comprehension module built around a meaningful educational task — for instance, writing a report on the scientific method integrating information from an encyclopedia entry, a newspaper article and a student lab report. These assessments seek to measure student performance against real-world tasks, rather than against a politically determined proficiency score, Gitomer said. “You’ve got this link to what it means to be competent,” Gitomer said. “You’re constantly helping the teacher and the student understand what that structure is that they’re really trying to move toward.”

The CBAL project faces challenges, Gitomer acknowledged. Equating the difficulty of different PAAs to ensure that results are comparable from year to year is complex. The tests must be computer-scored to keep costs down, but not every kind of task can be scored by computer. Nor does every school have the technology infrastructure to administer these kinds of tests, Gitomer said. Creating more complex and frequent assessments raises other practical questions, as well. “How are we going to pay for it all?” wondered **Lindsay A.L. Hunsicker**, a staff member to U.S. Republican Senator Michael B. Enzi of Wyoming.

More profoundly, Gitomer said, new assessments will catch on only if our political system abandons its current focus on a single proficiency score. “The hope in moving to a model like this is that it opens up the conversation,” Gitomer said. “If we just think about the achievement gap in terms of where kids are relative to a relatively low bar, I think we’ll have missed the point and be unsatisfied as a society, in terms of our international and national success and competitiveness.”

Assessing New Kinds of Skills

If CBAL seeks to test cognitive skills more effectively, the next frontier in testing may lie in assessing the noncognitive skills that influence success in college and the workplace — such qualities as persistence, integrity, leadership and motivation (see the graphic below for additional examples). Studies support the common-sense conclusion that these noncognitive variables are important to achievement in both school and the workplace, ETS researcher **Patrick C. Kyllonen** told the symposium audience. In one study, a researcher found that noncognitive factors predicted scores on an array of K – 12 achievement tests; another study found a similar impact on job performance and training time. “Both in education and in the workforce, we see that noncognitive skills are predicting outcomes,” Kyllonen said.

What Are the Noncognitive Skills?

Professionalism	Enthusiasm	Leading
Work ethic	Values	Deciding
Teamwork	Character	Supporting
Collaboration	Breadth	Cooperating
Oral communication	Open-mindedness	Interacting
Critical thinking	Persistence	Presenting
Agreeableness	Collegiality	Adapting
Ethics	Independence	Coping
Self-esteem	Motivation	Enterprising
Diversity	Planning	Performing
Leadership	Organization	Extroversion
Creativity	Self-efficacy	Emotional stability
Innovation	Anxiety	Conscientiousness
Lifelong learning	Self-concept	Openness

Source: ETS Center for New Constructs.

Research also suggests that noncognitive qualities are not immutable, Kyllonen said. A study based on scores on personality tests given at least a year apart found that some crucial noncognitive qualities change across the lifespan: emotional stability increases rapidly through childhood and early adulthood, reaching a plateau around age 37, for example, while openness to new experiences grows early in life, plateaus in middle age and

drops off in old age. “There’s a lot of stability in personality, but it’s not nearly as high as a lot of people have this conception of,” Kyllonen said. “Personality changes; it can be improved.” Research is examining how noncognitive skills, such as time management, can be improved and whether such improvements will yield corresponding improvements in student achievement, Kyllonen said.

Although it sounds innovative to educators, assessing such intangibles has long been common practice in industry, College Board Vice President **Wayne J. Camara** told the symposium audience. Through job analysis, employers identify desired outcomes, decide what qualities are necessary to achieve those outcomes, and find ways of measuring which job applicants possess those qualities. Applying similar methods in the college admissions process has the potential to yield significant benefits, Camara said. Today, colleges

rely heavily on admissions test scores and high school grades in deciding which applicants are likely to succeed, and these indicators do successfully predict freshman-year grades. But an industry-style job analysis of college success shows that it consists of much more than earning good grades; it also comprises returning to school each year, completing a degree and moving on to graduate training or satisfying work, Camara said. And these tasks demand a range of noncognitive qualities, from emotional stability to engagement with education, which colleges currently take into account only in their subjective, non-standardized admissions procedures. “We want a lot of behavior that transcends cognitive,” Camara said. “I would argue that we can measure these things reliably, fairly and objectively, and we don’t.”

‘We want a lot of behavior that transcends cognitive. I would argue that we can measure these things reliably, fairly and objectively, and we don’t.’

— Wayne Camara

Predictors of College Success

College Skills	Content Knowledge Achievement	Noncognitive	Personal Qualities/ Experiences/ Characteristics	School Performance/ Context	Interests – Vocations
Verbal Reasoning Math Reasoning Writing	Math Language Arts Science	Motivation Follow-through Communication	Letters Essay Community Service	Grades GPA Weighted GPA	Career Interests Learning – Study Skills Interest in Major
Metacognition Creativity Practical Knowledge Spatial Relations Intellectual Curiosity Technology – Research Skills	Social Studies/ Humanities Foreign Language Language Proficiency	Conscientiousness Leadership Other Personality Health/Lifestyle Adaptability	Extra-curricular Work Experience Literacy in Second Language Teacher Ratings Gender Ethnicity Ability to Benefit Family Education/ Income Ability to Pay Residence	Rank Courses Completed Academic Rigor AP/Honors Courses School Size School Quality	Self-efficacy Aspirations/ Realistic Self-concept

Tests measure
 Colleges collect in some form (applications, transcripts)
 Not collected in standard form

Source: Wayne J. Camara and Ernest W. Kimmel (Eds.), *Choosing Students: Higher Education Admissions Tools for the 21st Century*, Mahwah, N.J.: Lawrence Erlbaum Associates, 2005.

Kyllonen’s research assesses noncognitive skills using three criteria: student self-assessments, teacher ratings, and scores on tests of situational judgment, which ask test takers what they would do if, say, they had to organize a study group for students with conflicting schedules. Camara’s research uses both a situational judgment test and a “biodata” questionnaire, which asks respondents multiple-choice questions about their interests and past experiences. Researchers validated these measures on college juniors with respectable grades — the true experts about what success in college requires, Camara said — and then administered the same assessments to 3,300 freshmen at 11 colleges. The results of the noncognitive assessments contributed little to the prediction of freshman-year grades. “If you’re only interested in predicting grades in college, look no further than high school grades, SAT® and ACT®,” Camara said. But the results of the noncognitive assessments did significantly improve the prediction of other outcomes, such as graduation, absenteeism, leadership and engagement. A further study, still in progress, will administer the assessments to more than 11,000 applicants at 15 colleges and universities; these schools have agreed to follow enrolled students through their college careers to evaluate how well the noncognitive assessments predict performance on everything from grades and retention to absenteeism and institutional commitment. Any test items that appear biased — that predict the performance of women but not men, for instance, or of White but not African American students — will be discarded, Camara said.

Research suggests that using assessments of noncognitive ability in college admissions will produce a more diverse student body, Camara said, increasing the admittance rates of Hispanic and African American students, especially at the

most selective schools. Since these noncognitive assessments measure qualities that contribute to college success, it makes sense to find ways of incorporating them into the admissions process, he said. “We’re not talking about changing what we measure to increase diversity,” Camara said. “We’re talking about changing what we measure, and how we measure it, to make it more realistic to the environment, whether it’s college or whether it’s work.”

The Social Context

For education reformers, today’s state testing regime embodies a tension, symposium speakers made clear: Defining success according to a single proficiency score distorts the education system, but it also brings the achievement gap into focus. Standards fall short, curricula narrow, teachers lack diagnostic information — but, for the first time, Americans can see clearly the magnitude of school failure for low-income and minority children. Revamping the current testing system promises to yield richer information but risks sacrificing that clarity. “If we don’t have a quantifiable proficiency number that we’re shooting at for all kids,” said **Gary Huggins**, the director of the Aspen Institute’s Commission on No Child Left Behind, “how do we even identify the achievement gap and know what that is and do anything about it?”

‘If we don’t have a quantifiable proficiency number that we’re shooting at for all kids, how do we even identify the achievement gap and know what that is and do anything about it?’ — Gary Huggins

Implicit in Huggins’s question is a vision of what schools and tests can accomplish, a vision of a world in which policymakers force school improvement by holding educators accountable for closing the achievement gaps that tests reveal.

Missing from that vision — and, by design, from a symposium focused on the nitty-gritty work of improving standards and assessments — is the world outside the schoolhouse door. At a special session the night before the symposium, two speakers, economist and sociologist **Richard Rothstein** and ETS researcher **Paul Barton**, sought to place the problem of educational achievement gaps in a broader societal context.

The NCLB-inspired accountability system rests on a fundamental misconception about what it will take to close achievement gaps, said Rothstein, a research associate at the Economic Policy Institute (EPI). The roots of the problem lie not in the classroom but in the social conditions facing children who grow up in poverty. “Somehow, we continue to develop education policies in this country that expect schools alone to close the achievement gap, and No Child Left Behind is the latest iteration of that,” Rothstein said. “Clearly, expecting schools to wipe out the achievement gap on their own, without any support from the surrounding social environment, is something that’s bound to fail.”

‘Clearly, expecting schools to wipe out the achievement gap on their own, without any support from the surrounding social environment, is something that’s bound to fail.’ — Richard Rothstein

In 2003, Barton authored ETS’s *Parsing the Achievement Gap: Baselines for Tracking Progress*, a report that he said, “asked the question, ‘What gaps in life and school experience would have to be closed in order to close the achievement gap?’ ” Drawing on hundreds of studies, Barton identified 14 family, school and community factors — from low birth weight and lead exposure to class size and curricular rigor — that most researchers agree play a role in sustaining educational achievement gaps. On virtually all of these factors,

Barton found, gaps exist between the experiences of minority and non-minority children, and of low-income and higher-income families. Barton and ETS researcher Richard Coley are working on an update of the report, examining whether these gaps have narrowed in the past five years.

If non-school factors help create and sustain achievement gaps, it will take more than educational interventions to close them, argue the dozens of experts on education, health care and child welfare — including Rothstein and Barton — who signed a recent EPI statement calling for a “broader, bolder approach to education.” That new approach would require not only school improvement but also expansion of early childhood education, increased investment in health services, and the establishment of after-school and summer programs for low-income students.

The EPI statement’s message is not that schools do not matter or should not be held accountable, Rothstein stressed, nor that standardized testing has no part to play in assuring that accountability. But schools should be held accountable for what schools can do. “By holding them to impossible standards, we’re undermining their chances of improving,” he said. “We’re mislabeling schools as successful and failing if we expect them to achieve on their own what no school can achieve on its own.”

This emerging consensus, along with its implications for research and policy, was the focus of “Educational Testing in America: State Assessments, Achievement Gaps, National Policy and Innovations,” the 11th in ETS’s series of “Addressing Achievement Gaps” symposia, launched in 2004. The conference, cosponsored by the College Board, was held September 8 in Washington, D.C., and featured 13 researchers and policymakers as speakers, panelists and respondents. U.S. Secretary of Education **Margaret Spellings** gave the keynote address. Remarks were also delivered by Syracuse University Associate Vice President **Youlonda Copeland-Morgan**, the Chair-elect of the Board of Trustees of the College Board; ETS President and CEO **Kurt M. Landgraf**; ETS Senior Vice President **Michael T. Nettles**; and ETS Board of Trustees Chair **Piedad F. Robertson**. Sessions were moderated by Robertson and by ETS Senior Vice President **Ida Lawrence**; Morgan State University President **Earl S. Richardson**, an ETS trustee; and College Board Vice President **Ronald A. Williams**.

Symposium sessions included:

- State Assessments Today: What State Are We In?
- Assessment, Learning, Equity: What Will It Take to Move to the Next Level?
- Classroom Assessment FOR Learning and the Achievement Gap
- Redesigning K – 12 Assessment Systems: Implications for Theory, Implementation and Policy
- Lessons Learned from Industry: Achieving Diversity and Efficacy in College Success
- Enhancing Noncognitive Skills to Boost Academic Achievement

Supporting materials from the presentations are available as downloadable PDF or PowerPoint files at <http://www.ets.org/stateassessments>.

ETS Policy Notes is published by the
ETS Policy Information Center
Educational Testing Service
Rosedale Road, MS 19-R
Princeton, NJ 08541-0001
(609) 734-5212

e-mail: pic@ets.org

www.ets.org/research/pic

Director: Richard J. Coley

This article was written
by Deborah Yaffe
and edited by Richard J. Coley
and Richard Pliskin.

Copyright © 2009 by Educational
Testing Service. All rights reserved.

Educational Testing Service is
an Affirmative Action/Equal
Opportunity Employer.

ETS, the ETS logo and LISTENING.
LEARNING. LEADING. are registered
trademarks of Educational Testing
Service (ETS). SAT is a registered
trademark of the College Board. ACT
is a registered trademark of ACT, Inc.
10339

NON-PROFIT
ORGANIZATION
US POSTAGE
PAID
EDUCATIONAL
TESTING SERVICE



Listening. Learning. Leading.®

www.ets.org