Listening. Learning. Leading.®

2008

Research Portfolio
Research & Development Division

Research Conducted by ETS R&D in 2008

This document describes the breadth of the research being conducted in 2008 by the Research & Development division at ETS.

The research described below falls into three large categories:

(1) Research supported by the ETS research allocation,

(2) Research funded by testing programs at ETS, and

(3) Research funded by external governmental and private agencies.

Within each category, information is provided about specific research projects active this year, the focus and purpose of the research, and the R&D staff responsible for it. There is also discussion of why the research is important to do: how the work is aligned with ETS's mission and how it is building organizational knowledge and capability.

Section I: Research Funded in 2008 by the ETS Research Allocation

The ETS Research Allocation supports research at ETS that is aligned with the need to *innovate*.

In 2008, there are eleven initiatives supported by the Allocation:

1. Equating and Applied Psychometrics
2. Foundational Statistical and Psychometric Research
3. Psychometric Infrastructure
4. Cognitively Based Assessment of, for, and as Learning
5. New Constructs
6. English Language Learning and Assessment
7. Validity
8. Classroom Assessment and Practices
9. Constructed-Response Design and Scoring
10. Reinventing Test Development
11. Interactive Learning: Educational Applications

These initiatives fall into three categories, each aligned with a component of innovation: (1) pioneering research to create new knowledge and new capabilities, (2) using R&D knowledge and capabilities to maintain and enhance existing products to ensure ongoing quality and competitiveness, and (3) using R&D knowledge and capabilities to contribute to the development of new products. The work in these categories is interrelated. The new knowledge and capabilities generated in the first category are intended to help "feed" R&D's contributions to both new product development and the enhancement of existing products.

Below are brief descriptions of the work in 2008 in each initiative.

A. Initiatives that conduct pioneering research to create new knowledge and new capabilities

Name of initiative	Initiative description	Why is this research important to do?	ETS contacts
Foundational Statistical and Psychometric Research	This initiative is designed to develop and continuously improve upon the statistical and psychometric methodologies required to advance ETS's products and services. The primary focus of the initiative is on improving: • Continuous testing • Estimation and use of latent-variable models in testing applications • The quality of subscore information	--To ensure that the methodology used in psychometric operations is defensible and efficient in both a computational and statistical sense --To advance the field of educational measurement	Shelby Haberman & Matthias von Davier
Cognitively Based Assessments of, for, and as Learning	The central goal of this initiative is the creation of a future assessment system in reading, writing, and mathematics that takes a fundamentally different approach to K-12 school accountability and classroom testing. The approach attempts to synergistically unify three systems: accountability assessment, formative assessment, and professional support. The systems will built upon cognitive research, state standards, and curricular considerations. Work will include the: • Design of domain competency models (such a model specifies the knowledge, skills, and abilities important for success in a content domain and how these components are organized) • Creation and field testing of prototype tasks, assessment modules, and school-year assessment designs • Conducting of psychometric modeling of task and assessment performances within and across periodic accountability assessments • Development and adaptation of automated scoring models as appropriate for designated task models • Analysis and design of tools to report test information	--To solve a pressing educational problem (i.e., creating a balanced assessment system that gathers useful information for policy purposes and effectively supports classroom learning) --To advance the field of educational measurement by developing scientifically sound assessments that are considered by teachers to be educationally worthwhile	Randy Bennett & Drew Gitomer

B. Initiatives that use R&D knowledge and capabilities to maintain and enhance existing products to ensure ongoing quality and competitiveness

Name of initiative	Initiative description	Why is this research important to do?	ETS contacts
Equating and Applied Psychometrics	This initiative will develop applied psychometric and statistical methods and capabilities primarily focused on equating, linking, and quantitative fairness assessment. The initiative will: • Improve analysis related to assessment quality/equity, such as DIF analysis and examinee score equity indices • Support the operational implementation of the kernel equating (KE) method • Establish best practices for equating under suboptimal conditions • Improve current equating and linking practice (traditional equating, IRT equating, vertical scaling)	--To ensure the quality, equity, and fairness of assessments --To enhance the efficiency of testing programs --To advance the field of educational measurement	Alina von Davier
Psychometric Infrastructure	This initiative focuses on the development of statistical/psychometric infrastructure to increase operational and computational efficiency, allow more secure delivery of continuous testing, and prevent errors. The initiative will: • Identify methods for improving continuous testing • Standardize psychometric processes across programs and work groups • Continue enhancement of NAEP operational software, including both the operational systems used at ETS and DESI software • Add to or improve our general data processing hardware and software capabilities	--To increase confidence in the integrity and repeatability of the results produced from ETS's statistical systems --To improve operational and computational efficiency of ETS's software and data processing methods	Tim Davey
Reinventing Test Development	This initiative includes practical, applied investigations of new or revised systems, methodologies, or work processes that will lead to quality or efficiency improvements in the development of items, tests, and test-related materials. The initiative will: • Identify tools or approaches that will substantially increase the efficiency of test development work • Build prototype assessments	--To meet the demand for many more items and tests and do so in a more cost efficient way	Tom van Essen & Barbara Elkins

Name of initiative	Initiative description	Why is this research important to do?	ETS contacts
	• Develop guidelines for increasing accessibility of items and tests		
Validity	This initiative focuses on assuring technical quality for existing and new assessments for all individuals, including those with disabilities. The research in this initiative seeks to develop methodologies, provide guidelines, and build capacity at ETS to: • Support the psychometric quality of new and established ETS tests and products • Establish validity, fairness, and accessibility of assessments for students with disabilities • Expand standard-setting and job analysis methodology	--To meet ETS's mission to create assessments and assessment-related products that are fair for all learners --To respond to greater public demand for scientific evidence of the efficacy of ETS's assessment products and services	Brent Bridgeman
Constructed-Response Design and Scoring	The purpose of this initiative is twofold: • To gain a deeper understanding of current constructed-response scoring methodologies, practices, and policies relating to all content domains and to conduct research into improvements to the current practice of constructed-response scoring at ETS, both human and automated • To develop new prototypes and capabilities that address human and automated scoring in a variety of content domains, extended discourse structures, and simulation-based assessment responses	--To prevent error and increase efficiency by evaluating the current state of cross-program practices and formulating a strategy for constructed-response scoring across ETS programs	Catherine McClellan & David Williamson
English Language Learning and Assessment	• The purpose of this initiative is to develop assessments and related tools for English language learning and teaching. The Initiative is focused on improving our ability to build fair and valid assessments of both content area knowledge and English language proficiency for English Language Learners (ELLs). The initiative will: • Develop the scientific knowledge and related capabilities to improve existing assessments or develop new ones for English language learners • Develop the scientific knowledge and related capabilities to improve existing or develop new integrated assessment and learning systems for English language learners	--To advance ETS's mission to promote learning and educational performance for all people worldwide	John Young, Mary Enright & Maurice Hauck

C. Initiatives that use R&D knowledge and capabilities to contribute to the development of new products

Name of initiative	Initiative description	Why is this research important to do?	ETS contacts
New Constructs	The purpose of this initiative is to explore the feasibility of new constructs—both cognitive (e.g., critical thinking, communication skills) and noncognitive (e.g., work ethic, teamwork, leadership, ethics and integrity, and adaptability)—as the basis for new products and services that ETS could offer in the future. The initiative in 2008 will also focus on research to evaluate the validity of test scores and threats to validity. An additional focus will be on how to improve, as well as measure, noncognitive skills.	--To determine the importance of noncognitive skills for achievement --To develop new product concepts that measure constructs currently not being assessed	Patrick Kyllonen
Classroom Assessment and Practices	This initiative focuses on generation of knowledge and capability to improve teacher effectiveness, with a particular focus on classroom assessment and professional development. In 2008 the initiative will focus on: • Gathering data on teacher effectiveness as a way to inform potential products across several ETS business units. • Gathering additional data on the implementation and efficacy of the research-based teacher professional development program, Keeping Learning on Track (KLT) for the K-12 market.	--To advance ETS's mission by enabling educators to teach more effectively so that their students will succeed in school and in life	Cynthia Tocci
Interactive Learning: Educational Applications	This initiative focuses on research to support the design and development of engaging content for the family market (i.e., home-based products) that adheres to ETS values or quality and fairness. The content is intended to support learning for pre-secondary school children as well as their parents.	--To expand the base for our quality products and services by leveraging ETS's assessment and design methodologies and capabilities for the consumer market --To support the ETS mission to create products and services that promote learning and performance	Marisa Farnum

Section II: Research Funded in 2008 by ETS Testing Programs

In addition to research funded by internal research allocation, a number of research studies and other research activities are funded by the testing programs at ETS. These research studies and activities continue to ensure an adequate research base for the assessments offered by the different programs.

Some of the testing programs for which research is being carried out in 2008 are listed below. Next to each program is the name of the research liaison. The research liaison monitors research studies and ensures that they are completed on time and within budget and that all studies receive adequate technical reviews before their public release. These individuals also serve in the role of a high-level technical consultant for the program and, as appropriate, attend client and policy boards, advisory committees, and conferences representing the research, psychometric, and development concerns of the program.

Program	Liaison name
Test of English as a Foreign Language™ (TOEFL®)	Yasuyo Sawaki
Test of English for International Communication™ (TOEIC®)	Donald Powers
College Board Programs: SAT®, Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT®), Advanced Placement Program® (AP®)	Brent Bridgeman
Graduate Record Examinations® (GRE®)	Brent Bridgeman
PRAXIS™	Richard Tannenbaum
iSkills™	Irv Katz
e-Sir	Lydia Liu
Southern Regional Education Board (SREB)	John Young
Major Field Tests (MFT)	Guangming Ling
Measure of Academic Proficiency and Progress (MAPP)	Lydia Liu
K-12 Testing Programs: Comprehensive English Language Learning Assessment (CELLA), Oklahoma Modified Assessment, Miami-Dade Interim Assessment	Brent Bridgeman
Texas Teacher Certification Test Battery	Richard Tannenbaum

In addition, staff from the Psychometrics area of R&D continue to provide psychometric support for all of the testing programs. Each testing program also has a psychometrics manager assigned to it. Similar to the role of the research liaison, the psychometrics manager works with program staff to identify and prioritize program-specific psychometric needs and work as well as to monitor the work. The psychometrics manager ensures that each program maintains the highest psychometric quality possible.

The psychometrics managers for 2008 are

Program	Psychometrics Manager
College Board Programs AP/CLEP SAT/Subject/PSAT/NMSQT	 Michael Walker Jinghua Liu
Higher Ed Programs GRE PRAXIS Texas MAPP/iSkills/NBPTS MFT/Global MFT	 Fred Robin Kevin Larkin Fred McHale Michael Walker Jinghua Liu
K-12 Programs Chicago/CSU programs/ HSTW/Miami-Dade/Qatar STAR/CST/Standards Tests in Spanish/CMA/CAPA Maryland/Tennessee/Texas/ CELLA	 Venessa Lall Kevin Meara Jerry Gorham
Global and Workforce TOEFL/TOEIC ISkills Other	 Lin Wang Brad Moulder Alina VonDavier

A. Research studies in 2008

The bulk of program-funded research studies falls into five categories:

1. Providing validity evidence
2. Evaluating issues related to scores and scales
3. Using technology in support of scoring
4. Score interpretation
5. Understanding a test's psychometric properties

Validity evidence
Research that provides evidence in support of the intended inferences and actions based on the reported results for a testing program provide validity support for the test. Developing a validity rationale for a test and gathering the appropriate evidence is of importance here. The type of evidence gathered depends on the nature of the test, its scores, and the intended use of the scores. Research studies concerned with validity evidence may examine the test's construct representation, relationship to internal and external measures, ability of the test score to predict future performance, test content (job analyses, alignment with standards, etc.), and the impact of test use.

Some of the validity work to be done in 2008 by ETS R&D staff involves efforts to build a validity rationale for the testing program. Examples of this activity is the work that began in 2006 for the High Schools That Work (HSTW) for the Southern Regional Education Board, where a technical development plan and long-term research agenda was created, will be implemented in 2008; the development of a

construct validity argument for the PRAXIS series; and the creation of web-based documentation for TOEFL iBT quality and validity. In addition, research staff will provide consultative services on test validity for programs. For example, consultation will be provided to the CUNY campuses using iSkills to help them design, analyze, and interpret results of school-based validity studies; similar work is planned for College Board programs.

Other studies will work on providing specific types of validity evidence for a test. For example, a series of job analyses are planned which will serve to substantiate the content validity of the PRAXIS assessments and the Texas tests. Another study will look at content knowledge and TOEFL iBT reading performance. Other work that will be completed this year includes examining the predictive validity of the TOEFL iBT scores using faculty ratings and other criteria; the development of prediction models for the Chicago math assessments (CPS) and for the Miami-Dade tests; understanding the impact of student motivation on the construct validity of the Major Field Tests; investigating gender bias in student evaluation of teaching for e-SIR; and evaluating the impact of read-aloud accommodations on performance on a state-level English Language Arts test. A new study is underway that will look at value-added issues for the MAPP tests as part of the voluntary system of accountability for undergraduate schools. Finally, work on partnering with departments of education from several states as a way of systematically gathering data for the GRE program will continue with the expectation that data will be obtained from at least one state during 2008.

Scores and scales

Several studies that evaluate new scale issues, investigate whether an existing scale can be maintained, or examine the technical appropriateness of different scores (subscores, diagnostics, etc.) will be carried out this year. The development of technical manuals for the California Modified Assessment (CMA) and the California Standardized Testing (CST) program is planned and will document various scores and scales issues for these tests; an exploration of methods that could be used for creating a total score on the revised GRE is proposed; evaluating the impact of context effects on the scoring and equating of the revised GRE is underway; and scaling efforts related to the new Maryland modified high school assessment are planned. Other examples include investigating various methodological approaches to determine the value of section scores for the TOEFL iBT and completing a report on the reliability of subscores for MFT. Three studies concerned with aligning scores on one test with those of another tests are also planned: (1) comparing performance on the GRE and GMAT is continuing from 2007; (2) aligning scores from the Texas tests with those from PRAXIS tests is planned; and (3) investigating TOEFL and IELTS score alignment has begun.

Scoring and technology

During 2008 there will be continuing efforts related to the use of technology in supporting scoring for testing programs. In 2007, a large amount of work was done to evaluate e-rater for operational use in the TOEFL and GRE programs. This work will continue in 2008 with a focus on underlying validity issues for implementing e-rater in both programs. Efforts related to scoring speech samples will also continue in 2008. A study on the influence of rater language background on the reliability of speaking scores was begun in 2006 and will be finalized this year. This large study also has practical implications because it is evaluating the possibility of expanding the rater pool for TOEFL by using overseas raters. In addition, new models for rater use are of interest this year. For example, a study examining the feasibility of using one- versus two-raters in scoring of essays from the California State University EPT is proposed.

Score interpretation
A number of operational studies lend themselves to evaluating what a score recipient understands about the score as well as understanding the use made of the scores. A good deal of this work centers around the use of standard setting methods to create cut-scores that are valid and interpretable. In 2008, standard setting studies are planned for the HSTW, PRAXIS, TOEIC, Texas programs, and for a number of K-12 contracts (e.g., CAPA, CMA). Other activities will involve the creation of cut-scores for use with populations that extend beyond the traditional test population. Such studies include setting cut-scores on TOEFL for use with international teachers who teach in Kentucky and for use with international nurses.

Work on providing information to test-takers on measures of non-cognitive and personal skills is also underway. Since these measures do not lend themselves to a single or multiple scores, methods for providing appropriate feedback and suggestions for student improvement are being investigated.

Psychometric properties
Investigations that examine the psychometric characteristics that affect quality and validity at the item- and test-level will be ongoing during 2008. These operational studies include determining reliability estimates and the measurement error (e.g., standard error of measurement, standard error of difference, and conditional standard error of measurement), measures of item functioning (e.g., item difficulty, DIF), and speededness indicators. These characteristics are examined and routinely documented in the test analysis reports produced by the Statistical Analysis staff. In 2008, studies will investigate a wide range of psychometric issues, such as examining issues related to moving from formula-scoring to rights-only scoring for Advanced Placement tests; investigating methods for equating test scores for small samples; evaluating score change information for the PSAT/NMSQT; investigating improvements to SAT smoothing; and investigating various anchor test configurations. Some studies will be focused on particular scaling issues, such as comparing different IRT models in scaling and equating TOEFL iBT; examining scale stability for TOEIC Bridge; evaluating scale drift in a frequent test administration context; evaluating scale stability for TOEIC test-takers from Taiwan; and comparing various scaling methods for Advanced Placement.

B. Research capabilities development

Over the few years, efforts were focused on the development of research capabilities that could provide opportunities for programs to process test-taker data more efficiently. During 2008, a number of programs have sponsored specific studies aimed at evaluating the use of these capabilities with an expectation of moving them closer to operational implementation. For example, work is continuing to evaluate e-rater for operational use in both the TOEFL and GRE programs. Significant progress was made in 2007 on comparisons to scores from human raters under a number of scenarios and uses; studies in 2008 will focus on validity issues, especially those related to various ethnic/racial, gender, and language subgroups. Piloting the use of e-rater with other assessments, such as the work being done with the National Institute of Information Technology (NIIT), is underway and will be completed this year.

Work related to the deployment of other automated capabilities will also continue in 2008. For example, an investigation of SpeechRater accuracy for use in TOEFL Accelerator is underway. In addition, the use of automated capabilities to enhance internal test development processes is being proposed. The proposed study will develop a natural language processing tool for controlling verbal overlap on the operational version of the revised GRE.

Section III: Research Funded in 2008 by External Agencies

In addition to research funded by the Research Allocation and research funded by testing programs, ETS R&D will be engaged in 2008 in more than 20 individual research projects that have been funded by external agencies. Among the funding agencies are various divisions of the U.S. Department of Education (for example, the Institute of Education Sciences or the Office of Special Education Programs), the U.S. Department of Health and Human Services, and other governmental and private organizations.

In making decisions about what externally funded research to take on, ETS R&D is motivated by several factors, including the alignment of the research to ETS's mission, the opportunity to develop new knowledge and capabilities that can be leveraged for future work, and the match of staff expertise to the requirements of the research. In addition R&D is aware that our external research activities can have positive effect on our business.

Externally funded research allows ETS scientists to develop relationships with other practitioners which will sometimes lead to additional opportunities. One focus in 2008 will be finding further opportunities to partner with other researchers, especially those affiliated with universities. ETS R&D has recently established a Center for External Research. The goal of the Center is increase the amount of externally funded research conducted at ETS while making sure that that research is aligned with ETS mission and business interests.

Several key themes or directions organize most of the 2008 research projects active as of this writing. These include:

- Research in the Area of Early Childhood Education
- Research in the Area of Large-Scale Assessment
- Research in Adult Literacy
- Research in Reading
- Research on Teaching and Learning
- Research on Noncognitive Abilities

In the sections that follow, the research studies in these six areas are described in more detail, with information about their major activities, deliverables for 2008, key questions being investigated, and rationales for why ETS is participating in the studies.

A. Research in the area of early childhood education

ETS R&D staff members are participating in two research projects that are part of the Early Childhood Longitudinal Study (ECLS), a multiyear longitudinal study sponsored by the National Center for Educational Statistics, a center within the U.S. Department of Education's Institute of Education Sciences. The overall purpose of ECLS is to examine the effects of a number of family, school, community, and individual variables on children's development, early learning, and early performance in school. ECLS tracks two overlapping cohorts: a birth cohort and a kindergarten cohort. The birth cohort strand of the study follows a sample of children from nine months of age through kindergarten. The kindergarten cohort strand follows a sample from kindergarten through the 5th grade. Recently, the kindergarten cohort strand was extended to study the sample though the 8th grade.

ETS is also using items from the ECLS-K first and third grade assessments in mathematics to create second grade assessments.

ETS's staff members are also participating in a research project that is part of the Head Start and Childhood Experiences Study (FACES). FACES provides longitudinal information about the characteristics, experiences, and outcomes for children and families served by Head Start.

Finally, ETS R&D is helping to evaluate the effectiveness of the Big Math for Little Kids (BMLK) Program.

ETS's role in ECLS-related research is primarily to provide outcomes assessments. ETS's role in FACES is to provide information about the psychometric properties of assessments designed by ETS for other purposes—including assessments designed as part of ETS's ECLS research—to inform a policy decision about the possible use of these assessments in FACES

What Are Some Key Research Questions these Studies Are Attempting to Answer in 2008?

ECLS-B, and ECLS-K: How can we best describe and understand children's early development; their health care, nutrition, and physical well-being; their preparation for school; key transitions during the early childhood years; children's experiences in early care and education programs and at the beginning of kindergarten; and how do their early experiences relate to their later development, learning, and experiences in school?

FACES: What are the factors that influence the success of Head Start programs? What classroom quality and learning environments, curricular approaches, and teacher qualifications lead to the most successful outcomes in terms of school readiness, developmental gains, and changes in participant characteristics?

Evaluation of Mathematics Curricula: What are the best ways to teach Mathematics?

Analysis of Preschool and Kindergarten Assessment Data-EDC: How well does BMLK work?

Why Is It Important for ETS to Participate in this Research?

Connection to our mission: ECLS, FACES, and BMLK are all ultimately aimed at helping researchers identify the multiple factors with greatest impact on the academic success of young children. ETS's participation in these studies—as a provider of valid and reliable outcomes measures—is consistent with our mission to advance quality and equity in education.

Knowledge/capability building: Participation in these longitudinal projects provides ETS with valuable experience in providing accurate measurement of changes over time, a capability that can be leveraged for other uses, including building potential new assessments in response to No Child Left Behind (NCLB) legislation.

Who is the ETS R&D contact for this work?

Michelle Najarian

B. Research in the Area of Large-scale Assessment: Supporting International Assessments of Achievement

Overview

ETS R&D is doing research in 2008 involving large-scale assessments: that is, research to provide policy makers with information on what *populations* can and cannot do. The research is in support of international assessments of achievement.

PIAAC

In early 2008, ETS received word that the ETS-led consortium of seven international organizations (ETS, Westat, Deutsches Institut fur Internationale Padagogische Forschung (DIPF), cApStAn, IEA, Gesis-Zuma, and the Universiteit Maastricht) were selected to conduct PIAAC (Programme International for the Assessment of Adult Competencies) for the Organization for Economic and Cultural Development (OECD). Irwin Kirsch will lead the five-year effort to conduct one of the most ambitious and significant international surveys of human capital ever undertaken. The winning proposal was for an adaptive instrument that will be delivered by laptop computer in households in up to 35 countries. Instruments will measure four essential domains: Literacy, Reading Components, Numeracy, and Problem Solving in Technology Rich Environments.

PIAAC recognizes the supreme importance of human capital in influencing the social, educational, and economic outcomes of individuals and societies. The ETS work on PIAAC will provide policy makers worldwide with data that will influence policy decisions for many years to come. As of this writing (early March 2008), the details of the contract have yet to be worked out, so it is impossible to give detailed time lines and deliverables, but we expect that the work will commence in earnest in the 2nd quarter of 2008.

Other Research in Support of International Assessments of Achievement

IERI

In 2008, ETS R&D is in the second year of a multi-year collaboration with the International Association for the Evaluation of Educational Achievement (IEA), headquartered in the Netherlands. The two organizations have established a joint research and training unit—the IEA/ETS Research Institute (IERI) — with the mission of a) advancing and improving the science of international large-scale assessment, b) training and development of staff, and c) disseminating research findings. IEA is a member of the consortium that will work on PIAAC.

The activities of advancing and improving assessment will focus on carrying out a program of research. Research activities associated with these issues will be aimed at:

- developing improved test design and scaling methodologies as well as new methods to study relationships between proficiency data and other variables;
- developing and validating non-cognitive constructs that hold promise to predict cognitive measures on the level of policy-relevant groups;
- developing data collection methodologies that improve quality;
- addressing issues around international assessments that will help ensure innovation and incremental improvement of these assessment programs over time.

The training activities will be focused on the organizing, scheduling, and coordinating of training on specialized topics in large-scale assessment. The training will involve inviting renowned experts in the field to conduct multi-day training seminars. Possible topics will include, but are not limited to, item response theory, hierarchical linear modeling, and sampling.

Dissemination activities will focus on the launching the "Large-Scale Assessment Monograph Series." Each monograph will contain a series of papers related to the science of large-scale assessment.

TIMSS

Trends in International Mathematics and Science Study (TIMSS) measures students' progress in mathematics and science achievement on a regular four-year cycle, and permits reliable comparisons of the achievement of U.S. students with those in other countries. It collects educational achievement data at the 4th and 8th grades in approximately 50 countries to provide information about trends in performance over time together with extensive background information to address concerns about the quantity, quality, and content of instruction. ETS R&D is providing consulting services on TIMSS.

What Are Some Key Research Questions PIAAC and TIMSS Are Attempting to Answer?

- What is the state of human capital development in those countries participating in PIAAC?
- How well prepared are those surveyed to succeed in a 21st-century work environment?
- Are cognitive diagnosis models appropriate tools for reporting in international large scale assessments?
- Do omit rates vary across countries, and if yes, can differences in background data account for these differences?
- What are the important reading skills that should be assessed for students reaching the end of compulsory education? How to US students compare with other students in terms of knowledge of science and mathematics?

Why Is It Important for ETS to Participate in this Research?

Connection to our mission: The important international assessment comparisons and trends among countries will be used by researchers and policy makers to target resources and education interventions to advance quality and equity in education worldwide.

Knowledge/capability building: Participation in these studies provides ETS with an opportunity to contribute to the development and application of assessment methodologies in international settings. In addition to providing an opportunity to develop additional ETS expertise and experience in scaling international assessments, participation in these studies provides ETS the opportunity to build on relationships with international educators and agencies.

Who are the ETS R&D contacts for this work?

Irwin Kirsch and Tom Van Essen

C. Research on Adult Literacy

ETS R&D has a program of research, mainly using large-scale assessments and surveys, that measures the literacy (for example, the ability to work with prose or documents), numeracy, or computer literacy of adults, in both the United States and abroad. The research outside the United States involves adults in developed countries but also in developing ones, where there is much less of an existing assessment infrastructure. The frameworks and instruments to be developed for the international research studies need to reflect broad linguistic and cultural diversity. Because the international assessments and surveys focus on comparative outcomes, issues surrounding translation and adaptation take on importance.

Many of the assessments in this research program use or extend literacy measures that have been developed at ETS.

One of the studies being carried out in 2008 is funded by Statistics Canada and is looking at adult literacy and life skills. A second is funded by Princeton University and is looking at the motivation of 12th-grade students taking the NAEP reading assessment.

What Are Some Key Research Questions these Studies Individually Are Attempting to Answer?

Statistics Canada-funded study*:* What is the relationship between literacy skills and the economic, social, and personal characteristics of individuals and of nations?

Princeton NAEP study*:*

1. Do 12th grade students taking the NAEP reading assessment who are offered performance incentives display greater levels of engagement and/or achieve higher scaled scores on average than comparable students who are not offered such incentives?
2. Are there engagement or performance differences by treatment condition among students classified by ability level, by gender or by race/ethnicity?
3. If there are differences in performance by treatment group, what is the likely impact on the statistics reported by NAEP, as well as on other indicators that are constructed from NAEP data?
4. Are there detectable differences between the control condition (fall administration) and the standard NAEP spring administration?

Why Is It Important for ETS to Participate in this Research?

Connection to our mission: The important information about literacy levels that will emerge from these studies—in many cases for the first time—will be used by researchers and policy makers to target resources and educational interventions to improve adult literacy. ETS participation in this research, then, will help to advance quality and equity in education.

Knowledge/capability building: Participation in these studies provides ETS with an opportunity to contribute to the development and application of methodologies used to study growth and change. They also extend ETS's methodological capabilities in the translation and adaptation of assessment materials. In the case of Princeton NAEP, this work will provide information for policy makers who are considering an extension of NAEP.

Who is the ETS R&D contact for this work?

Irwin Kirsch

D. Research on Reading

Overview

This research direction includes a series of six related studies that together seek to identify the cognitive, linguistic, and neurobiological characteristics of struggling readers, the specific skills they lack, and some promising approaches for improving their abilities. These studies seek to identify the developmental course and prevalence of learning disabilities and how these disabilities interact with reading fluency, vocabulary and other oral language abilities, and comprehension skills.

The conceptual framework of these studies includes an assumption that the *components* of reading (for example, phonemic awareness, phonics knowledge, vocabulary knowledge, fluency, etc.) can be identified and that instruction can be targeted toward improving these components and integrating them into general reading comprehension ability.

Most of these studies make use of the Study of Adult Reading Acquisition (SARA), an experimental, computerized, component skills test battery. These studies seek to investigate the battery's use as a diagnostic tool, as a measure of student progress, and as an instrument for evaluating the relative effectiveness of instructional programs that differentially target specific reading components. In the case of one of the studies, the battery will serve as the foundation for the development of a new series of assessments to identify sources of reading comprehension difficulty.

Two of the six studies employ MRI brain imaging to understand the neurobiological characteristics that are associated with reading difficulties.

Four of the six studies focus on struggling adolescent readers; one focuses on struggling adult readers; and one focuses on both adolescent and adult readers.

In addition to these six related studies, there are two other studies that are part of this strand of reading research.

One study seeks to develop research-based principles and guidelines to make large-scale reading assessments more accessible for 4th and 8th grade students who have disabilities, while maintaining standards of validity. As do the four reading studies just described, this study is also experimenting with a component approach to enhancing the diagnostic potential of reading proficiency assessments. A related study is focused on improving state reading assessments for students with visual impairments. This study includes an investigation of the psychometric properties of state reading assessments for students with visual impairments as well as research and development of a prototype assessment of technology-assisted reading for such students.

What Are Some Key Research Questions these Studies Are Individually Attempting to Answer?

- What are sources of difficulty for readers in texts?
- What are promising programs in reading instruction for adults and adolescents?
- Is there a relationship between differences in cognitive, linguistic, and neurobiological characteristics that are associated with reading development and intervention outcomes?
  What is the prevalence of different types of reading disability among 4th and 8th grade learners?
- How can large-scale assessments be made more accessible for students who have learning disabilities and visual impairments that affect their reading?
- How do we define reading proficiency for students who use technology (screen readers, text-to-speech readers, refreshable Braille display) to read?
- How effective is an explicit literacy curriculum in improving the English reading, writing, and speaking skills of low-literate English as a second language (ESL) learners?
- What technologies are most effective in assisting students with visual impairments?
- How can the reading abilities of students with visual impairments best be assessed?

Why Is It Important for ETS to Participate in this Research?

Connection to our mission: Research to understand which interventions can improve the performance of struggling readers, to develop finer-grained classroom diagnostic and monitoring assessments, to identify how to make reading assessments more accessible to children and adolescents with disabilities, to accurately assess teacher knowledge of reading instruction, and to evaluate a curriculum aimed at improving the skills of adult English language learners, is directly connected to ETS's mission to advance quality and equity in education. The populations served by these studies are predominantly ethnic minorities, English language learners, and individuals with disabilities. Collectively, their reading performance illustrates the reading achievement gap in the United States.

Knowledge/capability building: Several of these studies allow ETS to gain experience in aspects of intervention study implementation. ETS is also able to learn more about component measures such as the SARA battery in order to make it applicable in a variety of school-based settings. Further, computationally identifying sources of text difficulty will assist assessment developers in the design of targeted classroom assessments. Some of these studies also allow us to better understand how to make assessments more accessible for English Language learners and for learners with disabilities.

Who are the ETS R&D contacts for this work?

John Sabatini, Jane Shore, and Cara Cahalan Laitusis

E. Research on Teaching and Learning

Overview

The studies in this category of research focus on improving student learning and achievement, improving teacher quality, and understanding the effect of psychology on cultural understanding.

Developing and Using Diagnostic Items in Mathematics and Science

The study developed a set of multiple-choice items that cover the major content areas in 4th and 8th grade mathematics and science. The incorrect answer choices in the items directly connect to student misconceptions. The investigators have piloted the items, along with strategies for using them in classrooms to improve student learning, on a group of 48 teachers In 2008 we will continue to administer and collect data from these items.

Collaborative for Middle-schools Mathematics and Science Project

This project shares characteristics of with the previous study. Quarterly assessments are being developed, along with analysis of student errors and misconceptions, and strategies will be developed to assist teachers in using the information from the assessments to inform their teaching. In addition, ETS staff are providing support for school districts as they develop teachers' understanding and practice of formative assessment approaches.

Both of these projects have a central focus on *formative assessment*. Formative assessment can be described as frequent, interactive assessment of student progress and understanding in which evidence of learning is collected and then used as feedback to adjust teaching and learning.

The Relationship between Mathematics Teachers' Content Knowledge and Students' Mathematics Achievement

This project examines the predictive relationship between middle school teachers' mathematics

knowledge—as measured by the teachers' scores on the *Praxis Series*™ Middle School Mathematics test—and students' achievement, as measured by their score gains on state-mandated mathematics tests. In essence, the study examines the contribution of teacher knowledge, as measured by a content licensure test, to student achievement.

Evaluating Transfer Learning in College-level Physics

This project aims to evaluate how effectively machines "learn." Specifically, the study will evaluate transfer learning in the context of college-level physics. *Transfer learning* involves using concepts and strategies learned through experience with one kind of problem in potentially novel ways to solve different kinds of problems. Transfer learning is a powerful form of learning, potentially capable of the kind of flexible, adaptive, and creative problem-solving that is the hallmark of human intelligence.

Using Assessment to Help Students Succeed is a two part study.

In Part 1 the investigators will produce a white paper on the feasibility of creating a comprehensive psychosocial (or, non-academic) assessment of college readiness for secondary students. The white paper will provide:

- a comprehensive framework identifying the key psychosocial factors related to school success;
- empirical evidence for the relationship between the different psychosocial factors and various academic outcomes, such as school grades, standardized test scores, and staying in school;
- suggestions for how psychosocial factors could be measured in an operational assessment system (methods would include self assessments, ratings by others—e.g., teachers and principals, and situational judgment tests).

The paper will make the case that psychosocial factors are important, that the educational community knows how to measure them, and that we can improve educational achievement, particularly for underserved students, by doing so.

Part 2 is a project to further the development of the *Keeping Learning on Track*™ (KLT) program—a research-based professional development program for teachers. American education needs improvement at any number of levels. One primary level is on finding ways to help teachers become even more effective. KLT does this by showing teachers how to integrate minute-to-minute, day-by-day formative assessment into their everyday teaching. The big idea that unifies and motivates KLT is the concept of "using evidence of learning to adapt instruction in real time to meet students' immediate learning needs." Evidence of learning is obtained through a series of powerful assessment techniques and modalities that allow teachers to make targeted interventions in order to make a real difference in real time. A pilot study of these interventions will be conducted with mathematics and science teachers in challenged high schools.

What Are Some Key Research Questions these Studies Are Individually Attempting to Answer?

- What is the contribution to learning of diagnostic feedback compared to other feedback conditions, such as simply accuracy information? Does providing teachers with diagnostic questions, suggestions for how to interpret and act on students responses, and associated professional development activities increase student achievement?
- Does an initiative that includes coaches, professional development, and alignment of curricula improve student learning in middle school mathematics and science? What is the contribution of teacher knowledge, as measured by the Praxis Series Middle School Mathematics Test, to students' middle-school mathematics achievement?

- What are the most effective strategies and study regimens for learning college-level physics? Can a "deep structure" analysis of a content area such as physics provide us with a basis for automatically generating assessment items in that area?
- How can non-cognitive assessments improve the likelihood of minority youth attending college?
- How can certain teacher strategies improve school performance in at risk neighborhoods?

Why Is It Important for ETS to Participate in this Research?

Connection to our mission: Research to improve student learning in science and/or mathematics, including for students with disabilities, is directly connected to ETS's mission to advance quality and equity in education.

Knowledge/capability building: A number of these studies will expand our understanding of both what formative assessment is and how teachers can be supported in developing this aspect of their instructional practice. The *Evaluating Transfer Learning in College-level Physics* study builds ETS's capability in the assessment of transfer learning, which may have important implications for redefining how achievement is assessed at the K-12 level. This study also gives ETS additional experience in using ETS-owned tools: Math Test Creation Assistant and Model Creator.

The study using the Praxis Middle School Mathematics Test will provide predictive validity information about the test, which could enhance its value. In addition, the study will shed light on the extent to which teacher's content knowledge directly impacts student achievement and will have implications for teacher certification requirements, preparation, and development. Finally, because it will build and apply an enhanced value-added model that explicitly accounts for covariate information and attenuation of teacher effects over time, this study will also contribute to the value-added research base.

The first part of the *Using Assessment to Help Students Succeed* project may very well establish a new paradigm for college admissions testing in this country.

In addition to these five research studies, ETS is participating in the National Comprehensive Center for Teacher Quality (NCCTQ). The U.S. Department of Education has funded a set of Regional Comprehensive Assistance Centers. Each center has a specific geographical region to serve. In addition, five comprehensive content centers provide support to the regional centers in a specific content area. The mission of the National Comprehensive Center for Teacher Quality (NCCTQ), in which ETS participates, is to serve as the premier national resource to which the Regional Comprehensive Assistance Centers, states, and other education stakeholders turn for strengthening the quality of teaching. The work of the NCCTQ is focused especially on issues related to high-poverty, low-performing, and hard-to-staff schools; the NCCTQ is also concerned with providing guidance in ensuring that highly qualified teachers are serving students with special needs.

Specific goals of the NCCTQ include:

- Promoting successful implementation of the teacher quality requirements of NCLB by disseminating critically reviewed research, strategies, practices, and tools
- Galvanizing public and policy maker support to meet the demands of NCLB related to teacher quality

The Department of Education funds the NCCTQ. ETS is a subcontractor to Learning Point Associates in the operation of the NCCTQ. The Education Commission of the States and Vanderbilt University are also partners in this effort.

ETS R&D staff have primary responsibility for producing the NCCTQ's research syntheses (two per year); research columns for the center's other publications and web site; and the *Biennial Report on Teacher Quality*. ETS staff will also organize and operate the higher education (teacher education) segment of the center's activities. Team members will contribute to the development of online resources and databases; conferences and network events; national advisory panel activities; and the general management of the center.

Because this center is operated under a cooperative agreement, the Department of Education will be closely involved in every aspect of the center's work, including selection of the topics to be covered by the research syntheses.

What Are Some Key Research Questions ETS Staff Are Attempting to Answer through their Involvement in the NCCTQ?

- What kind of school is difficult to staff?
- How can teacher retention be improved in such schools?
- How can teacher quality be raised in all schools?
- What effective materials are available to help schools meet the teacher quality requirements of NCLB?

Why Is It Important for ETS to Participate in this Research?

Connection to our mission: ETS's officers and trustees have made an organizational commitment to reducing achievement gaps. The lack of highly qualified teachers has been demonstrated to be a critical obstacle to reducing achievement gaps. The NCCTQ is directly charged with identifying and disseminating information about policies and practices that have been demonstrated to raise the quality of the teaching force, particularly for teachers of those placed at risk of academic failure through poverty and other factors. ETS's expertise in this area, and its power to reach key educational constituents, are strong assets in this work. The work with this center will demonstrate ETS's commitment to improving education and providing constructive, research-based solutions to meeting the requirements of NCLB.

Knowledge/capability building: ETS will become more knowledgeable and up to date on research and best practices in the critical area of teacher quality. We will also be learning about the intersection of improving teacher quality and meeting the needs of special education students. Finally, we will expand our corporate knowledge of how to administer federal content centers (as distinct from regional and R&D centers).

Who are the ETS R&D contacts for this work?

Caroline Wylie *(Developing and Using Diagnostic Items in Mathematics and Science; Collaborative for Middle-schools Mathematics and Science Project)*

Richard Tannenbaum *(The Relationship between Mathematics Teachers' Content Knowledge and Students' Mathematics Achievement)*

Patrick Kyllonen *(Evaluating Transfer Learning in College-level Physics)*

Tom Van Essen *(Using Assessment to Help Students Succeed)*

Laura Goe *(NCCTQ)*

F. Research on Noncognitive Abilities

ETS R&D has recently received funding for two studies to look at various aspects of non-cognitive ability. The first of these, *Multimedia Assessment of Emotional Abilities: Development and Validation,* will explore ways of measuring emotional abilities (EA, also referred to as emotional intelligence). In this project, computer-delivered tests of EA will be constructed and given to samples of community college and university students, along with measures of personality, ability, and outcomes. The methodologies used to develop the EA tests include: (1) the situational judgment test (SJT) paradigm (where participants rate a scenario for emotional relevance and/or salience); (2) an emotional principal-agent paradigm (EPAP, where event-emotion contingencies in others have to be perceived and memorized and emotion-behavior contingencies inferred from observed behavior to predict future behavior); (3) a cloze technique (where an emotional term completes, for example, a quote made by a famous philosopher); (4) various information-processing paradigms (with emotions as stimuli and speed of response as the variable of interest); and (5) an implicit association technique (where individual's implicit association of emotions with words and situations are assessed). These measures will then be evaluated in terms of their psychometric properties.

The second is *Psychological Dimensions of Cross-cultural Differences.* The objectives of this study are to (a) identify the psychological measures of personality, attitudes, and values known to be sensitive to cultural influences, (b) determine the dimensionality of the space defined by these measures, (c) examine the differences between cultural and national groups on these dimensions, and (d) define world regions on the basis of psychological dimensions. Through questionnaires and rating scales, personality, social attitudes, values and social norms known to show differences between cultures will be assessed.

Why Is It Important for ETS to Participate in this Research?

Connection to our mission: Central to the ETS mission is the sense of using assessment to help people. Non- cognitive measures are the next frontier in measurement science and it is incumbent upon ETS to explore all possible responsible uses for this kind of measurement..

Knowledge/capability building: Multimedia assessments of EA will have relevance for a number of kinds of assessment in the future. The technical problems solved during the creation of these tasks will be of use in the next generation of both cognitive and non cognitive tests. In addition, establishing a solid empirical basis of test of EA will support the development of noncognitive tests in other domains.

Who are the ETS R&D contacts for this work?

Richard Roberts and Lazar Stankov

Listening. Learning. Leading.®

www.ets.org