



Listening. Learning. Leading.®

2010 – 11

Research Portfolio

Research & Development Division

Copyright © 2010 by Educational Testing Service. All rights reserved. ETS, the ETS logo, LISTENING. LEARNING. LEADING., E-RATER, GRE, KEEPING LEARNING ON TRACK, TOEFL and TOEIC are registered trademarks of Educational Testing Service (ETS). EXADEP, KLT, PRAXIS, and TOEFL IBT are trademarks of ETS. SPEEHRATER is a service mark of ETS. All other trademarks are property of their respective owners.

Research ETS R&D is conducting in 2010-11

This document describes the breadth of the research that the ETS Research & Development division is conducting in 2010. This portfolio will be updated in early 2011 to reflect changes to existing projects and new projects that were added after this document was completed.

The research described in this portfolio falls into three large categories:

- (1) Research supported by the ETS Strategic Research Investment;
- (2) Research funded by testing programs at ETS; and
- (3) Research funded by external governmental and private agencies.

Within each category, this portfolio provides information about active research projects, the focus and purpose of the research, and the R&D staff members responsible for it. There is also discussion of why the research is important — how it is aligned with ETS's mission and how it builds organizational knowledge and capability.

Section I:

Initiatives With Funding from the ETS Strategic Research Investment

The Investment supports research at ETS that is aligned with the need to *innovate*. In 2010, the Investment is supporting 11 initiatives:

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
7. Validity
8. Understanding and Enhancing Teaching
9. Constructed-Response Design and Scoring
10. Test Development Quality, Efficiency and Innovation
11. Technology in Assessment

This list of initiatives will be updated in early 2011.

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 ETS's existing assessments in order to ensure their ongoing quality; and (3) using R&D knowledge and capabilities to contribute to the development of new assessments. 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. Brief descriptions of each 2010 initiative follow.

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

Initiative	Description	Why is this research important?	ETS contacts
Foundational Statistical and Psychometric Research	<p>This initiative aims to develop and continuously improve upon statistical and psychometric methodologies required to advance ETS's products and services. The initiative primarily focuses on improving:</p> <ul style="list-style-type: none"> • Continuous testing • Estimation and use of latent-variable models in testing applications • The quality of subscore information 	<ul style="list-style-type: none"> • 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
Technology in Assessment	<p>This initiative aims to explore the use of technology to fundamentally improve assessment—from item design through delivery, scoring and reporting. In 2010 the initiative will focus on:</p> <ul style="list-style-type: none"> • Test delivery on user-supplied hardware • Applications of multiprocessor and cloud computing 	<ul style="list-style-type: none"> • To build foundational knowledge to eventually increase: <ul style="list-style-type: none"> a) productivity (by improving test development and scoring) b) quality (by developing new ways of delivering items that gather more accurate and richer information) 	Mike Wagner

B. Initiatives that use R&D knowledge and capabilities to maintain and enhance ETS's existing assessments to ensure their ongoing quality

Initiative	Description	Why is this research important?	ETS contacts
Equating and Applied Psychometrics	<p>This initiative aims to develop and apply psychometric and statistical methods and capabilities that are necessary to ensure the equity and fairness—and to improve the efficiency—of ongoing testing programs. The initiative's primary focus is on equating. However, research on other applied methodologies is also included. The initiative will:</p> <ul style="list-style-type: none"> • Identify methodologies to ensure the stability and meaning of reported scores over time • Improve current equating and linking methodology (comparing equating methods, evaluating smoothing methods, and studying effects of sampling on equating) • Improve analysis related to subscore reporting • Improve communication of analyses results to clients 	<ul style="list-style-type: none"> • To ensure the quality, equity, and fairness of assessments • To enhance the efficiency of testing programs • To help clients better understand how analyses results should be interpreted 	Alina von Davier & James Carlson

Initiative	Description	Why is this research important?	ETS contacts
Psychometric Infrastructure	<p>This initiative focuses on the development of statistical/psychometric infrastructure to increase operational and computational efficiency, and prevent errors. The initiative will:</p> <ul style="list-style-type: none"> • 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 	<ul style="list-style-type: none"> • 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
Test Development Quality and Efficiency, and Innovation	<p>This initiative targets tools and methods for developing high-quality, innovative assessments in more efficient ways. The initiative will:</p> <ul style="list-style-type: none"> • Identify tools or approaches that will substantially increase the efficiency of test development work • Investigate ways of increasing the form-to-form comparability of assessments • Identify prototype innovative task types 	<ul style="list-style-type: none"> • To meet the demand for many more items and tests and do so in a more cost efficient way • To better measure traditional constructs and assess constructs that have previously been unmeasured 	Tom van Essen & Luis Saldivia

Initiative	Description	Why is this research important?	ETS contacts
Validity	<p>This initiative focuses on assuring technical quality for existing and new assessments for all individuals. The research in this initiative seeks to develop methodologies, provide guidelines, and build capacity at ETS to:</p> <ul style="list-style-type: none"> • 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 • Improve quantitative procedures used at ETS to help ensure fair assessment, including differential item functioning, and differential prediction 	<ul style="list-style-type: none"> • 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	<p>This initiative aims to develop new prototypes and capabilities that address automated scoring of constructed response tasks. The focus is on automated scoring of:</p> <ul style="list-style-type: none"> • Content (the meaning of ideas and the logical relationships among ideas expressed in text entered on a computer); • Speech (spoken responses of English language learners) 	<ul style="list-style-type: none"> • To facilitate better construct representation by allowing for the inclusion of constructed-response tasks where it was previously infeasible to administer and score such tasks • To reduce the cost and effort of scoring constructed-response items 	David Williamson

Initiative	Description	Why is this research important?	ETS contacts
English Language Learning	<p>This initiative aims to create foundational knowledge about assessing English Language Learners (ELLs), improve existing assessments for ELLs, and develop new assessment frameworks and prototype assessments for ELLs. In 2010, the initiative will focus on:</p> <ul style="list-style-type: none"> • English language proficiency assessments for domestic and international K-12 and post-secondary purposes • English language issues related to K-12 content area assessments in the United States 	To advance ETS’s mission to promote learning and educational performance for all people worldwide	Xiaoming Xi & Mikyung Wolf

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

Name of initiative	Initiative description	Why is this research important?	ETS contacts
New Constructs	<p>This initiative aims to explore the feasibility of noncognitive constructs such as work ethic, teamwork, leadership, ethics and integrity, and adaptability as the basis for new products and services that ETS could offer in the future. In 2010, the initiative will focus on:</p> <ul style="list-style-type: none"> • Improving the way background factors are assessed in questionnaires on international survey assessments • Identifying important noncognitive skills, attitudes, and dispositions for teachers, and investigating how these can be improved. 	<ul style="list-style-type: none"> • To expand our ability to assess, in valid and reliable ways, constructs other than cognitive ability and academic achievement • To investigate ways to improve noncognitive skill levels 	Patrick Kyllonen
Understanding and Enhancing Teaching	<p>This initiative focuses on the generation of knowledge and capability to build improved measures of teacher effectiveness. In 2010, the initiative will focus on:</p> <ul style="list-style-type: none"> • Measuring teachers' content knowledge (in mathematics, English/language arts, and science) <i>for</i> teaching • Investigating measures of teacher effectiveness, including the validity of subject-specific and general observation protocols 	<ul style="list-style-type: none"> • To advance ETS's mission by enabling educators to teach more effectively so that their students will succeed in school and in life 	Drew Gitomer

Name of initiative	Initiative description	Why is this research important?	ETS contacts
Cognitively Based Assessments of, for, and as Learning	<p>This initiative's central goal 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 build upon cognitive research, state standards, and curricular considerations. Work will include:</p> <ul style="list-style-type: none"> • Design of domain competency models (such a model that 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 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 	<ul style="list-style-type: none"> • 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

Section II:

Research With Funding From ETS Testing Programs

In addition to research funded by the Strategic Research Investment, R&D performs a number of research studies and other activities funded by the testing programs at ETS. These studies and activities define appropriate and specific work for a particular testing program and support the program by ensuring the technical and psychometric quality of the assessment.

Listed below are some of the testing programs for which R&D is conducting research in 2010. This list will be updated in early 2011. Next to each program is the name of the research liaison. Research liaisons serve as high-level technical consultants for the program and, as appropriate, attend client and policy boards, advisory committees, and conferences representing the research concerns of the program. Research liaisons also work with testing programs to identify research needs, determine appropriate funding sources, develop program specific research agendas, and monitor studies to ensure before public release that they are completed on time, within budget, and according to technical standards.

Program	Research Liaison
College Board Programs	Brent Bridgeman
GRE® General and Subject Tests (GRE)	Brent Bridgeman
High Schools That Work (HSTW)	John Young
K-12 Programs	John Young
Proficiency Profile	Lydia Liu
Major Field Tests (MFT)	Guangming Ling
PRAXIS™	Richard Tannenbaum
TOEFL®	Xioming Xi and Donald Powers
TOEIC®	Donald Powers

In addition, R&D's staff provides psychometric support for all testing programs. Each testing program is assigned a psychometric manager. In roles similar to those of the research liaisons, psychometric managers work with program staff to identify and prioritize program-specific psychometric needs, monitor the quality of work, serve as high-level psychometric consultants for the program and, as appropriate, attend client, technical and policy boards, advisory committees, and conferences representing the psychometric and development concerns of the program. The psychometrics manager ensures that each program maintains the highest possible psychometric quality. The psychometrics managers for 2010 are:

Program	Psychometric Manager
<i>College Board Programs</i>	
AP®/CLEP/SAT Subject Tests	Michael Walker
SAT/PSAT®/NMSQT®	Jinghua Liu
<i>Higher Ed Programs</i>	
GRE General Test	Fred Robin
PRAXIS	Kevin Larkin
Texas	Gautam Puhan
Proficiency Profile/MFT/GRE Subjects Tests	Michael Walker
<i>K-12 Programs</i>	
HSTW/MGA/CBAL/ERB / Tennessee/ Maryland / Chicago /CSU/EAP/Texas	Christine Mills
CAHSEE/CAPA/CST/ CMA/STS/WCAP	Valerie Link
<i>Global and Workforce</i>	
TOEFL	Venessa Lall
ALTS/EXADEP™/TOEIC/Qatar	Brad Moulder

The research liaison and the psychometric manager assigned to each program work together to develop studies and other research activities that support the program. The research studies and activities are aligned within a common framework, which produces consistency in level and type of work done in support of testing programs, aids testing programs in meeting ETS audit standards, assists in the prioritization of needed research on a program level, and determines the fundamental work needed for revising a program. This framework captures the major areas of program-based research and ensures validity, fairness, and quality for testing programs. The framework also allows for more efficient knowledge sharing across programs. The framework consists of 11 categories of research:

- Providing validity evidence
- Evaluating fairness and accessibility issues
- Supporting test revisions and ongoing program changes
- Evaluating issues related to scores and scales
- Investigating security concerns
- Applying technology in the support of test scoring
- Evaluating the Impact of test preparation
- Understanding candidates and populations
- Understanding a test's psychometric properties
- Evaluating score interpretation
- Policy issues in test use

Focus of Research

At this time, the majority of program-funded research studies fall into five of the framework categories:

- (1) Providing validity evidence
- (2) Evaluating issues related to scores and scales
- (3) Understanding the meaning of scores
- (4) Understanding the test's psychometric properties
- (5) Using technology in support of scoring

Providing Validity Evidence

Research provides validity evidence for a test when it provides evidence in support of the intended inferences and actions to be made based on the reported test results. Developing a validity rationale for a test and gathering the appropriate evidence is important. 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 (concurrent validity), ability of the test score to predict future performance for admissions and placement decisions, test content (job analyses, alignment with standards, etc.), dimensionality of the test, and the consequential impact associated with the use of the resulting scores.

Some of the validity work ETS R&D staff members are conducting in 2010 involves efforts to build validity rationales for testing programs. Several studies will provide validity evidence for admissions and placement in courses or programs. For example:

- Admissions and placement validity studies will be conducted for requesting institutions and participants in the College Board's national SAT validity study through the Admitted Class Evaluation Service (ACES) program.
- Course placement studies are ongoing for the California State University English Placement Test (EPM) and Entry Level Mathematics Test (ELM).
- In addition, a validity study will continue to investigate how useful TOEFL Internet-based Test (iBT) scores are for placement decisions in ESL programs.

A number of studies are more predictive in nature, including:

- a study that will use teacher ratings as an alternative to GPA when evaluating the relationship of TOEFL iBT scores to university success;
- an evaluation of how well TOEFL iBT scores predict students' academic English skills and academic success in college;
- the development of prediction models for the Chicago Public Schools (CPS) math and reading assessments;
- and a 3-year study for the GRE program will continue on the predictive validity of the ETS Personal Potential Index (PPI).

Several research activities will focus on additional types of validity evidence for the High Schools That Work (HSTW) program, including a concurrent validity study, a report on the history of HSTW, and a review of validity studies conducted to date. In addition, R&D will conduct a study on the effect of a read-aloud modification on item characteristics from the California Standards Test (CST) program. Ongoing validity efforts for GRE that allow for more in-depth analyses of various issues related to graduate education will continue. To do this, ETS is partnering with Florida and Kentucky, states that maintain an extensive student record system that will allow the use of additional criteria as part of validity research.

Research activities conducted for the PRAXIS program include studies to provide validity evidence in support of an assessment of teacher leadership and to support the creation of attitude and personality assessments for teacher professional development. R&D will provide validity research consultation for iCritical Thinking (iCT) and coordinate research aimed at supporting valid score inferences for the Major Field Tests (MFT) program.

R&D will also conduct research on variables that impact test-taker performance, and thus, the validity of a test. Such studies will include:

- a comparison of the impact of laptop and desktop use on TOEFL iBT writing scores;
- an administration mode comparison study for the Washington Comprehensive Assessment Program (WCAP);
- and a study on the relationship between test preparation and TOEFL iBT performance.

Evaluating Issues Related to Scores and Scales

This year, R&D will carry out several studies that evaluate new scale issues, evaluate mode comparability, or examine the technical appropriateness of different scores (e.g., subscores, diagnostics). For example, in 2010 R&D plans to

- conduct research on a conversion to rights scoring for the Advanced Placement (AP) program;
- examine ways to ensure comparability of scores across forms and establish new scales for all AP subjects;
- evaluate and report descriptive statistics for the original and revised HSTW indices;
- monitor the scale score distribution for quality control for the College Level Examination Program (CLEP);
- develop new scales for the Tennessee End-of-Course (TN EOC) program;
- bridge the new 2010 Mathematics scale to the 2009 scale for WCAP.

In addition, R&D will conduct research on the impact that test- and test-taker population characteristics have on scores and scales, as well as on equating results and calibration. Such studies include investigations of:

- the impact of item scrambling on item difficulty and scoring tables for WCAP;
- the best approach to equating when there is a population shift for SAT Subject Tests;
- repeater effects in equating SAT subject tests;
- and the effect of a 3-month retesting interval on test performance for CLEP.

R&D will also conduct studies on methods to track, measure, and use longitudinal test scores. Specifically, R&D will conduct research on growth measures for California assessments and develop statistical software for use with growth-based applications used in K-12 testing programs. In addition, we have designed a series of four studies related to measuring value-added outcomes for higher education institutions using the Proficiency Profile.

Several reports will document the technical quality of ETS products and services to ensure the soundness of different scores. Specifically, R&D will develop or create:

- technical manuals or reports for every K-12 client assessment program;
- a manual of standard setting methodologies and procedures for the TOEIC test;
- a product development technical report for the Educational Records Bureau – Computer Adaptive Achievement Test (ERB-CAAT);
- and a series of web-based documents will be developed to document evidence of quality for TOEFL iBT.

Understanding the Meaning of Scores

Research that evaluates score recipients' understanding of scores ensures that intended inference from a test score is aligned with the actual use and meaning of a reported score. This year, R&D will examine ways of improving the utility of TOEFL score reports by understanding the needs of TOEFL score users. We will conduct standard setting studies to create cutscores that are valid and interpretable. Specifically, standard setting studies will determine the cutscore needed to obtain a passing score for several subjects of the end-of-course assessments for the Tennessee Comprehensive Assessment Program (TN EOC) and for WCAP. In addition, R&D will develop a standard setting manual for TOEIC score users, which identifies appropriate standard setting methodologies and procedures.

Efforts related to determining how students view the role of and use GRE scores in graduate admissions will also continue. While the focus of a study completed in 2009 was on students who had taken the GRE, the focus in 2010 will be on students who may have not yet considered applying to graduate school.

Understanding a Test's Psychometric Properties

In 2010, R&D will continue ongoing investigations of psychometric characteristics that impact quality and validity at the item- and test-level. Such operational research includes

- statistical analysis in support of operational testing for Chicago Public Schools (CPS) benchmark assessments;
- the impact of item exposure for CLEP;
- studies on speededness for CLEP and AP;
- an examination of the impact of using different criterion scores on identifying DIF items in the anchor set for TOEIC; and
- a study on the reliability of speaking and writing measures for TOEIC.

Other research on the psychometric properties of assessments includes studies on the appropriateness and impact of new equating approaches and models. Exemplary studies include:

- an evaluation of a new approach for equating TOEIC forms;
- a comparison of subscore equating methods for SAT;
- and a study on the appropriate proportion of native speakers taking the CLEP Spanish test that should be included for more reliable test equating results.

Applying technology in the Support of Test Scoring

In 2010, R&D staff will continue efforts related to the use of technology in support of scoring for testing programs. In 2008 and 2009, much work was done to evaluate the e-rater® engine for operational use in the TOEFL and GRE programs and to implement the e-rater application with the GRE and TOEFL programs. This work will continue in 2010 with a focus on:

- understanding demographic-based differences in mean essay scores between human and e-rater evaluation for GRE;
- exploring the use of paraphrasing in the TOEFL Integrated summary task to better represent the writing construct in e-rater;
- evaluating analytic-based content scoring techniques in e-rater for Integrated writing tasks on TOEFL;
- and investigating TOEFL criterion prompts for e-rater models.

Research to advance the automated scoring of speech using SpeechRaterSM will also continue in 2010.

Research Work that Drives Policy

Work on a comprehensive report examining the political, demographic, socioeconomic, educational, and financial trends that impact participation in and completion of graduate education will be completed this year. The assumption underlying the report is that the global competitiveness of the United States and capacity for innovation hinges fundamentally on a strong system of graduate education. Trends in enrollment, attrition, and time-to-degree completion rates as well as international challenges will be discussed. Key vulnerabilities and threats that exist in the current graduate education system, within government, and in industry will also be presented. Finally, the report will detail policy recommendations based on empirical findings. These recommendations include actions that universities, industry, and government must take in order to ensure continuation of a strong system of graduate education and, as a result, national prosperity.

Section III: Research With Funding From External Agencies

The Center for External Research in ETS R&D coordinates more than 35 externally-funded research projects¹ that ETS Research & Development scientists work on. 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, Westat, and The Bill & Melinda Gates Foundation.

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

Externally-funded research allows ETS scientists to develop relationships with other practitioners, develop and expand their expertise, and to expand the amount of mission-related research that we do.

Most externally-funded research projects fall into one of seven large categories. These are:

- Measures of literacy
- Noncognitive assessments
- Assessment of people with disabilities
- Measures of the global distribution of human capital
- Measures of teaching quality
- Assessments in support of learning
- Advanced psychometrics

In the sections that follow, we describe the goals of current projects within these areas and work that was underway in 2010. This portfolio will be updated in early 2011 to reflect projects that began after this document was released and to reflect changes to existing projects. For further information please contact the principal investigators listed with each project or Thomas Van Essen, the Executive Director of the Center for External Research (tvanessen@ets.org).

¹ Some of the smaller projects are not featured in this document.

Measures of Literacy

Finding ways to identify the cognitive, linguistic, and neurobiological characteristics as well as the skill levels of struggling readers is one of the key educational challenges of our time. Many of the studies described in this section 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.

IES/ETS Reading Assessment Conference: Assessing Reading in the 21st Century

Principal Investigator: John Sabatini

Funding Source: U.S. Department of Education's Institute of Education Sciences

In 2008, ETS organized a conference to examine questions on how to align advances in reading science and measurement with the applied assessment practices that govern school and policy-level decision making. The conference brought together groups of scholars in the research community to present on emerging research and models of reading and reading assessment. Since the conference, the principal investigator has been working to publish the presentations in an edited publication.

In 2010, project directors were planning these activities:

- Complete review book featuring the papers that contributors presented at the conference
- Promote the book and its findings at conferences and to interested parties

Prevalence of Reading Difficulties in the 4th through 9th Grades

Principal Investigator: John Sabatini

Funding Source: National Institute of Child Health and Human Development (ETS is a subcontractor to the Kennedy Krieger Institute)

ETS is a subcontractor to Kennedy Krieger Institute on this National Institute of Child Health and Human Development (NICHD) project. The central goal of this project is to estimate the prevalence of various types of reading disability (RD) in the population of upper elementary and middle school students (grades 4 and 8). Determining the prevalence and distribution of reading difficulties across this age range and within subgroups of the population is of critical importance for designing reading instruction and school services that can best address the needs of struggling readers beyond the primary grades.

In 2010, project directors were planning these activities:

- Recruit a stratified sample of more than 45 schools across the state of New Jersey
- Collect data on reading for more than 3,000 learners in 4th and 8th grades, including English language learners, using a battery of reading assessments
- Collect data from parents of learners participating in the study, including information on each child's family background, educational background, and other factors that may influence learning (data from Spanish speaking parents will be included)
- Begin analyses of data to better understand components of reading needed to identify reading difficulties, and their subtypes
 - This will include setting criteria for the identification of reading disabilities, criteria for classification of reading disability types, and comparing scores for all reading disability types.

The participating schools are stratified by enrollment size (large, medium and small) and by the socioeconomic and ethnic composition of the community the school serves. Prevalence estimates will be derived for overall reading disability and for each subtype of reading disability. It is anticipated that the reading disability profiles will include deficits, alone and in combination, in comprehending text, recognizing/decoding words, and reading fluency. Each reading disability subtype also will be examined for relative prevalence by demographic factors (age/grade, gender, socioeconomic status, and ethnicity) and other student characteristics, especially English language proficiency. By collecting data about family, community, school, and instructional contexts, ETS will be able to investigate relationships between these factors and reading disability as well as to identify potential risk and protective factors within these contexts.

SARA (Study Aid and Reading Assistant)

Principal Investigator: John Sabatini / Kelly Bruce

Funding Source: Multiple sources of funding

SARA is a suite of diagnostic assessments based on a componential understanding of reading. SARA is built upon a research base of studies from the fields of cognition, linguistics, and neuroscience. SARA is appropriate for readers across the lifespan and can be used with learners in a variety of settings, including middle and high school, community college, and adult literacy programs. SARA batteries are computer-delivered and modularized, giving the suite the flexibility to respond to the varying needs of students and teachers as well as researchers. To date, SARA batteries have been administered to adolescents and adults in seven states and five countries. SARA's focus is on the components of reading — the subskills that contribute to reading proficiency — which include (1) word recognition and decoding, (2) vocabulary and morphology, (3) sentence processing, (4) silent and oral reading fluency, and (5) text comprehension. The program has various sources of external funding, with SERP, Inc., (Strategic Education Research Partnership) being the most notable. SARA also is in use in the Programme for the International Assessment for Adult Competencies (PIAAC) and may be used in the Programme for International Student Assessment (PISA).

In 2010, project directors were planning these activities:

- Collect field test data on the spring (post-test) version of the SARA from at least 5 participating middle schools
- Analyze SARA post-test data and produce research reports on subtest and item performance
- Develop SARA battery for use in PISA field trial in 2011 and analyze field trial results (contract is pending)
- Analyze results from the PIAAC field trial of 2010
- Conduct SARA data collection with adult learning centers in Philadelphia, PA, and provide professional development to adult learning instructors around the use of assessment data
- Continue SARA task and item development, particularly in partnership with allied projects in vocabulary and text analysis

Noncognitive Assessments

For most of its history, ETS has focused on assessing traditional academic skills—verbal reasoning, mathematical reasoning, and subject-specific knowledge in various academic disciplines. In recent years, the Research & Development division has placed a good deal of emphasis on assessing the so-called “noncognitive skills” because of a growing awareness within the research community of the relationship that personality and related noncognitive factors have to educational outcomes.

Multimedia Assessment of Emotional Abilities

Principal Investigator: Richard Roberts

Funding Source: Army Research Institute

For this Army Research Institute (ARI) contract, ETS is developing assessment instruments according to a taxonomic model that evolved from recent research. Using a variety of both new and established methodologies, tests will be constructed and given to samples of community college and university students in the United States. The project is evaluating psychometric properties of these new measures as well as correlations with criteria like GPA, mental health, and perceived social support.

Research is ongoing. Findings thus far indicate that the multimedia assessments are face-valid and have acceptable measurement properties. Thus, Emotional Ability (EA) measures were shown to be reliable both internally and over time. Relationships with intelligence, other emotion measures, and personality were as might be expected in the light of current theory. Relationships with valued outcomes were also suggestive. For example, EA was found to be related to how happy participants reported being while doing different daily activities.

The approach used to assess satisfaction in this study is derived from an economic model and indicates that EA may play a role in worker (and by extension, soldier) satisfaction and productivity. Clearly, there is a need to conduct experimental studies (e.g., manipulating decision-making under stress and comparing people identified as high and low EA), but thus far scores from these assessments have demonstrable construct validity.

ETS has used knowledge gained from this work to develop a module of ETS's ReadyEdge product. For its part, ARI is exploring use of the research in its applied selection research programs. Measures of EA could be used to identify personnel in need of increased training in working to understand and better manage their own emotions and the emotions of others. The new EA assessments are expected to be free from the adverse impact on minorities seen with traditional cognitive measures. In addition, the measures may provide predictive information regarding behaviors such as decision-making under stress, leadership, and working in a team that cognitive measures do not address.

In 2010, project directors were planning these activities:

- Complete experimental study examining whether emotional abilities predict decision-making under stress (collaboration with University of Cincinnati)
- Complete multivariate correlational study examining measures of emotional abilities and indicators of leadership
- Prepare manuscripts for peer-review publication
- Prepare and complete report for client
- Present findings at major conferences

Assessment of People with Disabilities

As a mission-driven organization, ETS is concerned that all students receive fair and valid assessments. A significant strand of activity within the Research and Development Division is devoted to meeting the special assessment needs of people with disabilities.

Designing Accessible Reading Assessments (DARA)

Principal Investigator: Cara Cahalan Laitusis

Funding Source: U.S. Department of Education

The Designing Accessible Reading Assessments (DARA) project was funded by the USDOE in October 2004 and is one of three projects funded as part of the National Accessible Reading Assessment Projects (NARAP).

The DARA project is completing its program of research to support the design and development of an accessible diagnostic reading assessment for students with reading-based disabilities. The focus of the DARA research has been on a multi-stage test design that allows for the isolated measurement of the components of reading (decoding, fluency, vocabulary, and comprehension) when the student scores poorly on an integrated stage one reading comprehension assessment.

The research from the DARA project has included experimental design, psychometric analysis of operational test data, and more qualitative procedures such as focus groups and think-aloud studies. DARA project research was a key source for *Principles and Guidelines for Accessible Reading Assessments*, (www.narap.info/publications/reports/definingreadingprof.pdf), which NARAP issued in 2009.

In 2010, project directors were planning these activities:

- Complete a field test of a two-stage component-based accessible reading assessment to 1,200 eighth-grade students in Massachusetts

The two-stage component-based accessible reading assessment will also include measure of oral reading fluency. The test design has implications for state accountability assessments as well as NAEP and could serve as a model for an adaptive assessment in an accountability framework. Detailed information can be found at the project Web site www.ets.org/dara, as well as the NARAP Web site www.NARAP.info.

Technology Assisted Reading Assessment (TARA)

Principal Investigator: Cara Cahalan Laitusis

Funding Source: U.S. Department of Education

The U.S. Department of Education funded the Technology Assisted Reading Assessment (TARA) project in July 2006. TARA is one of three projects funded as part of the National Accessible Reading Assessment Projects (NARAP) and includes a program of research and development to improve reading assessments for students with visual impairments or blindness. TARA is examining the properties of existing assessments for students with visual impairments and is developing an assessment of reading with a particular focus on independent technology assisted reading. Supported by the Center for Applied Special Technology (CAST), ETS and the National Center on Educational Outcomes (NCEO) co-lead TARA.

TARA's program of research includes:

- Psychometric research, which uses data from operational state assessments to determine how effectively current reading and English language arts assessments perform for students with visual impairments or blindness
- Focus groups and observational interviews with assistive technology readers to determine the necessary skills required to be an independent and efficient assistive technology (AT) reader
- Survey of AT readers to define the population of students who would benefit from an assessment of technology assisted reading and to inform the test blueprint
- Usability studies to ensure the prototype assessments created for TARA and the other NARAP projects are accessible and valid for students with visual impairments and blindness

In 2010, project directors were planning these activities:

- Complete a field test of an on-demand performance assessment to 200 blind or low vision students across the United States

This assessment will measure a student's ability to independently access text (using assistive technology) and could be adapted to serve as a district or state accountability assessment. This test design model also could be employed to construct alternate assessments for tests that are difficult to render in an appropriate format for blind and low vision test takers (e.g., scenario-based technology enabled assessments). Detailed information and the first research reports can be found at the project website, www.naraptara.info, as well as the NARAP website, www.NARAP.info.

***Feedback-and-Revision on Alternate Assessments
Based on Modified Achievement Standards (AA-MAS) in Mathematics***

Principal Investigators: Cara Cahalan Laitusis / Yigal Atali

Funding Source: U.S. Department of Education

With the National Center on Educational Outcomes (NCEO), ETS is investigating an approach to testing K-12 students with disabilities in which students are offered a second chance to answer a multiple choice question if they get the question wrong on the first attempt. The study stems from state efforts to show that their alternate assessment based on modified achievement standards (AA-MAS) meet the requirements of the U.S. Department of Education's Title I and IDEA regulations. Peer review of this evidence found that none of the states met the requirements and that an area for improvement was the research and rationale in support of modifications to test design and test item manipulation.

ETS and NCEO are exploring a feedback-and-revision model in which students who respond incorrectly to a test item would be able to revise their answer for partial credit. In the process, one distracter (wrong answer choice) would be removed after the student attempts the test item and the student would receive immediate feedback before making the revision. Prior research on providing immediate feedback-and-revision opportunities during assessments indicates that this change increases the reliability of test scores, reduces test anxiety, and improves learning. These benefits are expected to be particularly beneficial for students with learning disabilities on state assessments.

Awarded in January 2010, the project will begin with a series of small studies exploring a variety of response and scoring formats, item types, and content areas using cognitive interview methodology. Based on the results of these studies, the project team will conduct a large study to examine the psychometric and affective impact of revising wrong answer choices for partial credit on the reliability and validity of accountability assessment scores. A broad view of validity will include psychometric comparability and the accuracy of scores. In addition, the project team will examine the perceptions of students and their teachers on the testing experience and usefulness of test scores for instructional decision making.

In 2010, project directors were planning these activities:

- Attain firm commitments from states to participate in the study
- Set up cognitive interviews, including selecting and uploading test items for computer-based delivery and determining a system for data collection
- Conduct first set of cognitive interviews and consider revisions to the methodology based on the response
- Conduct second and third sets of cognitive interviews

Measures of the Global Distribution of Human Capital

As is made abundantly clear in the ETS report *America's Perfect Storm* (www.ets.org/Media/Research/pdf/PICSTORM.pdf), the United States and the world are facing crisis of human capital. Success in the 21st century will require a highly educated workforce. Policy makers — at both the national and international level — recognize the supreme

importance of human capital in influencing the social, educational, and economic outcomes of individuals and societies.

An area of research related to the development of human capital is research into the effect of early childhood education. ETS Research & Development staff members support the Early Childhood Longitudinal Study (ECLS), a multi-year 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. ETS's role in ECLS-related research is primarily to provide outcomes assessments. In addition ETS provides information about the psychometric properties of other assessments of early childhood education.

Feasibility Study for an Assessment of Higher Education Learning Outcomes (AHELO)

Principal Investigator: Thomas VanEssen/Claire Melican

Funding Source: Organization for Economic Co-operation and Development (OECD) (ETS is a subrecipient to the Australian Council for Educational Research)

The OECD is sponsoring a feasibility study of its Assessment of Higher Education Learning Outcomes program. As part of a larger two-year effort funded in January 2010, ETS will be responsible for instrument development work for an economics assessment. The ETS development team is working with a group of economics experts to produce a framework that broadly reflects the current thinking in tertiary economics education. Guided by the framework, ETS will work with a group of national experts to develop an economics assessment that will be given to graduates majoring in the field in selected schools in five countries (Belgium [Flemish language], Italy, Mexico, the Netherlands, Russia, and the United States). The development team also will be involved in the field tests of the new AHELO economics test and the data analysis of the results. The team will prepare reports as needed and work with other contractors on the overall AHELO effort. If successful, the AHELO feasibility study will demonstrate that it is possible to provide meaningful international comparisons between institutions of higher education.

In 2010, project directors were planning these activities:

- Work with international expert group to develop a framework for economics
- Work with the international expert group to produce an economics assessment aligned with the framework
- Work with national project managers in Mexico, Russia, Italy, the Netherlands, and Belgium to prepare field work in 2011

Programme for the International Assessment for Adult Competencies (PIAAC)

Principal Investigator: Irwin Kirsch / Claudia Tamassia

Funding Source: Organization for Economic Co-operation and Development (OECD)

The OECD has contracted with ETS to manage the consortium responsible for the Programme for International Assessment of Adult Competencies (PIAAC). The five-year study will examine 21st-century skills of adults in more than two dozen countries. The assessment will take place across OECD and partner countries in 2011 with results published in 2013.

The survey assessment will measure the literacy and numeracy skills of people between the ages of 16 and 65. It also will look at how well these individuals solve problems in technology-rich environments. The OECD commissioned the assessment as a way of investigating the link between skills and outcomes in educational, social and labor-market contexts. Specifically, the OECD expects that PIAAC will accomplish several goals:

- Provide policymakers in each participating country with a profile of their country's adult population in terms of the knowledge, skills and competencies thought to underlie personal and societal success
- Assess the relationship between these competencies and various social, educational and economic outcomes
- Gauge how successful education and training systems are at generating these outcomes
- Help to identify the factors policymakers could address to enhance these competencies

The other organizations in the ETS-led consortium are: Westat, cApStAn, the University of Maastricht, Gesis-ZUMA, the German Institute for International Education Research and the International Association for the Evaluation of Educational Achievement.

In 2010, project directors were planning these activities:

- Complete a field test of the assessment, which will be administered in 41 languages in 27 countries
- Perform a quality control check of the operational features of the testing platform, including support procedures
- Confirm that resulting data is complimentary with other adult surveys, such as the International Adult Literacy Survey

Programme for International Student Assessment (PISA)

Principal Investigator: Patrick Kyllonen / Irwin Kirsch

Funding Source: Organization for Economic Co-operation and Development (ETS is a subrecipient to the Australian Council for Educational Research)

The Programme for International Student Assessment (PISA) is a project of OECD. It assesses the reading, mathematics and science literacy of 15-year-old students. Three literacy domains are assessed: Reading; Mathematics; and Science. PISA assesses how well students approaching the end of their compulsory education are prepared for life beyond the classroom by focusing on the application of knowledge and skills to problems with a real-life context. The aim of PISA is to provide information on the following questions:

- How well are young adults prepared to meet the challenges of the future?
- Are they able to analyze reason and communicate their ideas effectively?
- Do they have the capacity to continue learning throughout life?
- Are some kinds of teaching and school organization more effective than others?

PISA provides information for use by policymakers and researchers throughout the world. Leading international experts work to develop the assessments whose results are comparable across different national and cultural contexts. The survey is implemented in three-yearly intervals. Tests are typically administered to between 4,500 and 10,000 students in each country.

ETS is part of a consortium led by the Australian Council for Educational Research (ACER) and is primarily responsible for the development of the PISA 2012 background questionnaires for students and schools participating in the assessment. Major activities to be completed in 2010:

- Conduct cross-national cognitive labs of the new PISA 2012 background questionnaire items in the United States, Germany, Hong Kong, Taiwan, China, Mexico, and Australia in both paper-and-pencil and computerized form
- Develop new background questionnaires to assess student health, student time use, parental attitudes, and other factors (countries will have the option to include these questionnaires in their PISA administration)

U.S. Participation in the Program for International Assessment of Adult Competencies (PIAAC)

Principal Investigator: Irwin Kirsch

Funding Source: U.S. Department of Education (ETS is a subrecipient to Westat)

The U.S. Department of Education has contracted Westat to manage the U.S. participation in the Programme for International Assessment of Adult Competencies (PIAAC). ETS manages the international consortium for the program. As a subrecipient to Westat, ETS is primarily responsible for the analysis and reporting of the U.S. results from the assessment. ETS will produce the first draft of an in-country report for the United States (set for release in 2013), and the effort will use software ETS developed for use in the NAEP Data Explorer to render tables and charts for the report. In 2010, project directors were planning these activities: :

- Work with the project lead to draft and finalize a sampling plan for a field test of the effort
- Help develop a recruitment and data plan for data collectors

Test of International Math and Science (TIMSS)

Principal Investigator: John Barone / Ted Blew

Funding Source: U.S. Department of Education's Institute of Education Sciences

The Trends in International Mathematics and Science Study (TIMSS) is the largest and longest standing international comparative study of its kind to date. TIMSS 2007 is the fourth in a cycle of international comparative assessments dedicated to improving teaching and learning in mathematics and science for students around the world. TIMSS is carried out every four years at

the fourth and eighth grades and provides data about trends in mathematics and science achievement over time. These studies are undertaken under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), based in the Netherlands. The international coordination of TIMSS activities takes place out of Boston College.

ETS is a subcontractor to Boston College on this project and in 2010 will continue to offer technical support on the maintenance and use of the data set.

Early Childhood Longitudinal Study, ECLS-K:11

Principal Investigator: Michelle Najarian

Funding Source: Westat / U.S. Department of Education's Institute of Education Sciences

ETS is a subcontractor to Westat on this, the third in this series of longitudinal studies sponsored by the U.S. Department of Education's Institute of Education Sciences. Like its predecessors, the ECLS-K:11 provides a rich and comprehensive source of information on children's early learning and development, transitions into kindergarten and beyond, and progress through school for a new cohort of children. Coming more than a decade after the ECLS-K:98, it will allow cross-cohort comparisons of two nationally-representative kindergarten classes experiencing different policy, educational, and demographic environments. The study will follow a nationally representative cohort of children in kindergarten in the 2010-2011 school year.

In 2010, project directors were planning these activities:

- Develop test forms in reading, math, and science for kindergarten, first and second grades for the national data collections in 2010-2013
- Analyze test data from the fall kindergarten national assessment in reading and math

Head Start Family and Child Experiences Survey (FACES)

Principal Investigator: Michelle Najarian

Funding Source: Mathematica Policy Research / U.S. Department of Health and Human Services

Since its founding four decades ago, Head Start has served as the nation's premier federally funded early childhood intervention. Focusing on children in the years before formal schooling, often from families with multiple risks, it has provided a natural and national laboratory for a wide range of basic, prevention, early intervention, and program evaluation research. The Family and Child Experiences Survey (FACES), first launched in 1997 as a periodic, longitudinal study of program performance, remains Head Start's ongoing flagship research initiative. The study is designed to be a reliable source of data for describing the experiences of Head Start children and their families; the quality of Head Start classrooms; and the qualifications, credentials, and opinions of Head Start staff.

Mathematica's five-year study chronicles the 2006 cohort of approximately 3,500 three- and four-year-old children enrolled in 60 Head Start programs from around the country. Each cohort is followed from entrance into the Head Start program, through one or two years of program participation, with follow-up in the spring of kindergarten. Through direct child assessments in multiple domains, observations of Head Start classrooms, and interviews with Head Start parents, teachers, and administrators, FACES gathers comprehensive data on the cognitive and social-emotional development of Head Start children; the characteristics of their families; the

quality of Head Start classrooms; and the qualifications, credentials, and views of Head Start teachers and other program staff. ETS offers technical and data analysis support to this effort.

In 2010, project directors were planning these activities:

- Analyze test data from the fall 2009 and spring 2010 national assessments in math and letter recognition

Measures of Teaching Quality

Good teaching is one of the most important factors influencing a child's education. Efforts to improve teaching quality are central to ETS's mission and the focus of considerable effort in Research and Development. ETS established the Center for Teaching Quality under the leadership of Drew Gitomer to further this work. Important grants from the Bill & Melinda Gates Foundation and the William T. Grant/Spencer Foundations testify to the importance of the work being carried out there.

Changing Landscape for Prospective Teachers in Career and Technical Education

Principal Investigator: Drew Gitomer

Funding Source: U.S. Department of Education's Institute of Education Sciences

With the significant amount of policy activity targeted at improving teacher quality, ETS is examining the changing characteristics of prospective teachers pursuing certification in career and technical education fields. The study compares demographic and academic characteristics for these teachers across two cohorts. The study also compares these cohorts of prospective CTE teachers with other certification areas using findings from an earlier ETS study, *The Academic Quality of Prospective Teachers: The Impact of Admissions and Licensure Testing* (www.ets.org/Media/Research/pdf/RR-03-35.pdf). The new study includes prospective career and technical education teachers who took licensure tests in the following areas: Agriculture, Business Education, Family and Consumer Science, Health Education, Marketing Education, and Technology Education. Each of these tests has been given in multiple states across the two cohorts. The states to be included in the study are those that (a) required the assessments listed above for purposes of initial teacher certification in the cohorts of 1994 to 1997 and 2002 to 2005 and (b) that have a sufficient number of individuals who took the test in those years so that the state can be included in the analysis.

In 2010, project directors were planning these activities:

- Complete final report and issue an ETS research report on the results

National Comprehensive Center for Teacher Quality

Principal Investigator: Laura Goe

Funding Source: Learning Point Associates / U.S. Department of Education

ETS is part of a consortium led by Learning Point Associates (LPA) that has established the National Comprehensive Center for Teacher Quality. The center is funded through a five-year contract from the U.S. Department of Education and is part of the department's comprehensive center program that includes 16 regional centers and 5 national content centers. The mission of National Comprehensive Center for Teacher Quality is to ensure a highly qualified teacher in every classroom. In addition to ETS and Learning Point Associates, the partnership includes the

Education Commission of the States (ECS) and Vanderbilt University. Working together as the Center, they serve as a national resource for regional centers, state education agencies, and other educational stakeholders. The Center provides guidance for strengthening the quality of teaching—especially in high-poverty, low-performing, and hard-to-staff schools—as well as schools that serve students with special needs. The website, www.tqsource.org, provides interactive databases, comprehensive research studies, publications, and other resources.

In 2010, project directors were planning these activities:

- Continue to offer technical assistance to states
- Develop a work plan as requested by the U.S. Department of Education for an extension year
- Work with partners on a proposal for the next five-year cycle of funding support

Understanding Teacher Quality

Funding Source: Bill & Melinda Gates Foundation

Principal Investigators: Drew Gitomer / Courtney Bell

ETS is leading a collaboration that will study teaching evaluation systems to gain a fuller understanding of what makes effective practice. The study also will investigate how various measures can provide information to support teacher development over the course of their careers. ETS is working with the University of Michigan's Institute for Social Research and The RAND Corporation on the "Understanding Teaching Quality" project. The Bill & Melinda Gates Foundation supports the effort through its Measures of Effective Teaching (MET) venture (www.metproject.org/). The three-year project is evaluating the validity, utility and interrelationships of a range of measures currently in use to assess teaching effectiveness. The aim of the collaboration is to provide an empirical and methodological base for the development of robust teaching evaluation systems by the fall of 2012. The study is investigating the relationships of central teaching characteristics to student learning and engagement. These characteristics include qualities of instructional practice, classroom climate, and emotional support for students and teacher knowledge. For the study, the researchers are working with a group of middle school math and English language arts teachers over a two-year period. Each teacher is being observed four times over the course of one school year for the purpose of collecting evidence on a number of measures of teaching effectiveness. During these observational visits, the project team also is collecting data about student assignments and teacher knowledge. In 2010, project directors were planning these activities:

- Complete data collection for Year 1 and the first half of Year 2, including classroom observations, teacher knowledge assessments, classroom assignments, and Year 1 value-added estimates for teachers.
- Conduct initial analyses of the study's instruments and the relationships between instruments

Measures of Effective Teaching / Teacher Knowledge

Principal Investigators: Drew Gitomer & Courtney Bell

Funding Source: Bill & Melinda Gates Foundation

Researchers from ETS and the University of Michigan will design and develop assessments that measure teachers' knowledge of the specific subject matters they teach. The Bill & Melinda Gates Foundation supports the effort through its Measures of Effective Teaching (MET) venture (www.metproject.org/). The basis of new measures will be a framework that taps into the specific subject matter knowledge teachers regularly use in their day-to-day practice. The effort will produce a set of assessments by December 2011. Through MET, the researchers will develop frameworks for the assessment of teacher knowledge with the specific focus on content pedagogy and content knowledge. The project team will then use the frameworks to develop assessments of teacher knowledge in mathematics and English language arts in grades 4-9 that address both general knowledge and specific focal topics.

In 2010, project directors were planning these activities:

- Develop assessments of Knowledge for Teaching for English Language Arts and Mathematics
- Complete a pilot test of the Knowledge for Teaching Assessments along with Praxis™ content measures and compare results
- Produce a framework for the design and development of knowledge for teaching assessments

Toward an Understanding of Classroom Context

Principal Investigators: Drew Gitomer / Courtney Bell

Funding Source: William T. Grant and Spencer Foundations

ETS, the University of Virginia, and The RAND Corporation in 2008 were awarded a three-year grant to explore the best practices in implementing the secondary school version of the Classroom Assessment Scoring System (CLASS-S), a promising observational instrument that measures student-teacher interactions. The University of Virginia developed CLASS, and its use has been validated in more than 2,000 elementary school classrooms. The ETS-led effort will test the validity of the CLASS-S—an altered version of CLASS meant for use in secondary school classrooms. The research is assessing and refining different ways to implement CLASS to optimize its efficiency. The project is gathering data from eighth- and ninth-grade algebra classrooms using three different strategies for completing the CLASS: conventional observation by a classroom rater, video recording of classrooms to allow for offsite coding, and teacher self-assessment. For the validity analyses, the project is collecting measures of teacher and student characteristics, teachers' math knowledge for instruction and teachers' knowledge of instructional support. The team is comparing CLASS and other scores to the changes in student scores on a standardized algebra test.

In 2010, project directors were planning these activities:

- Work with recruited school researchers to implement the study design, make the necessary school visits, and code the classroom performances live and via video recordings
- Use the data collected to help refine the Classroom Assessment Scoring System
- Write technical and training manuals for the system
- Begin to implement the dissemination plan
- Continue to consult with teachers, principals, technical experts, and stakeholders at the National Comprehensive Center for Teacher Quality

A Technology-Rich Teacher Professional Development Intervention that Supports Content-based Curriculum Development for English Language Learners

Principal Investigator: Jill Burstein

Funding Source: U.S. Department of Education's Institute of Education Sciences, Educational Technology Program

The goal of this project is to develop a technology-rich teacher professional development package (TPD) to support in-service, content-area teachers in the preparation of accessible content materials to facilitate content comprehension, language skills and reading comprehension of ELLs. Through iterative pilot studies with in-service teachers enrolled in college level courses offered by Stanford University and George Washington University, ETS over the course of the study will develop and evaluate the feasibility of an online TPD package. This package includes two components: (1) *TPD component*, and (2) *Instructional Authoring component*. The *TPD component* will provide teachers with a rich set of teaching strategies targeting the diverse learning needs of English language learners. The *Instructional Authoring component* will allow teachers to apply linguistic insight gained in the TPD to pedagogically adapt content-area curriculum for English language learners. The *Instructional Authoring* tools will incorporate the use *Text Adaptor*, a Web-based tool developed by ETS to provide automated text analysis and support for efficiently adapting classroom texts and instructional materials to learner needs.

In 2010, project directors were planning these activities:

- Initiate development of the *Instructional Authoring component* tools, including iterative system evaluations that will inform the development of the final technology platform
- Begin integration of the TPD content into the George Washington University and Stanford University courses

Assessments in the Support of Learning

An important strand of ETS's Research & Development activity concerns various efforts to expand the usefulness of assessment—especially efforts to use assessment in ways that go beyond summative (measures of what students have done) to measures that provide useful information that will support learning.

Creation and Dissemination of Upper-elementary Mathematics Assessment Modules

Principal Investigator: Judy Hickman

Funding Source: Harvard Graduate School of Education / National Science Foundation

The project is part of the Harvard Graduate School of Education's Learning Mathematics for Teaching (LMT) effort. ETS has joined Harvard and University of Michigan Institute for Social Research on the project. To help provide better evidence regarding the effects of teacher learning on student outcomes, the project is designing and pilot testing growth-sensitive student assessment modules for the upper elementary grades. The modules have three distinct features. First, assessment tasks are being designed from the literature on student's cognitive growth in particular mathematical domains. The assessments are focused on specific areas (number and operations; geometry and measurement; pre-algebra) and tasks place students along a rough developmental spectrum to allow evaluators to gauge growth over the course of a school year. Second, the modules are aligned with focal points released by the National Council of Teachers of Mathematics and existing LMT teacher knowledge and classroom observation instruments. The latter provides the National Science Foundation's Math-Science Partnerships and others a suite of instruments to detect any effects of programs on improved teacher knowledge and student achievement. Finally, the instruments are machine-scoreable enabling their use in studies that must recruit a large numbers of students to ensure adequate power to detect program effects.

In 2010, project directors were planning these activities:

- Complete the spring field test in grades 4 and 5 in Florida schools of four test forms (N = 500) to get preliminary statistics on the items
- Use ETS's test development system to store items and statistics
- Prepare paper (specific to Geometry/Measurement) on Models of Mathematical Development
- Work with Geometry/Measurement Advisory Council to develop prototype items for new test modules (followed by piloting items)
- Design six test forms with vertical linking design
- Complete fall field test (followed by additional fields tests in the spring and fall of 2011) with students in grades 4 and 5 (N = 2,000)

Cumulative Learning Using Embedded Assessment Results (CLEAR)

Principal Investigator: Lydia Liu

Funding Source: University of California – Berkeley / National Science Foundation

This project is taking advantage of new technologies and research findings to investigate ways that science assessments can both capture and contribute to cumulative, integrated learning of key concepts in middle school courses. ETS is a subcontractor to the University of California – Berkeley on this project. The team is researching new forms of assessment that document students' accumulation of knowledge and also serve as learning events. Aligning assessment and instruction around the goal of promoting understanding can improve learning outcomes and make any science, technology, engineering and mathematics course more efficient.

The project team is designing pretests, posttests, annual, and embedded assessments. The scaffolded knowledge integration framework and the cumulative learning strategies are guiding the design of assessments and instruction. The team is tracking progress in knowledge integration as well as trajectories of student energy conceptions and misconceptions.

The CLEAR team is conducting cohort comparison studies and randomized experiments comparing alternative strategies for promoting coherence in four participating schools. The quasi-experimental (cohort comparison) and randomly assigned classroom comparison methods are being augmented with student interviews, video case studies, and classroom ethnographies. In 2010, project directors were planning these activities:

- Design annual post-test for all modules across the sixth and seventh grades
- Design new assessment types, such as the two-tier multiple-choice items, to capture student reasoning in an effective way
- Analyze data from the post-test and provide feedback to the CLEAR leadership team
- Disseminate research results at the National Association for Research in Science Teaching Conference

Visualizing to Integrate Science Understanding for All Learners (VISUAL)

Principal Investigator: Lydia Liu

Funding Source: University of California – Berkeley / National Science Foundation

ETS is a subcontractor to the University of California – Berkeley on this project. The project is a study of the effectiveness of interactive, online animations or visualizations—which include models, simulations, and virtual experiments—to affect instruction by making unseen scientific phenomena, such as chemical reactions, visible to learners. VISUAL aims to transform instruction in physical science by:

- Creating designs and assessments for visualizations that promote coherent understanding of physical science concepts,
- Developing visualization-based curriculum materials that broaden participation in physical science by motivating all learners to succeed, and
- Designing authorable, cyber-enabled instructional materials that implement successful practices and provide guidelines for future designers.

In 2010, project directors were planning these activities:

- Develop an assessment that will provide a baseline measure of students' science proficiency and can be used to compare to student performance after the instruction
- Validate the assessment after data collection, including looking at the psychometric properties of the items and their reliability
- Obtain feedback on the research plan through the project's advisory board meeting

Enhancing the Formative Assessment Capacity in Pueblo School District 60

Principal Investigator: Caroline Wylie

Funding Source: University of Colorado – Colorado Springs / U.S. Department of Education's Institute of Education Sciences

The University of Colorado – Colorado Springs is trying to enhance the formative assessment capacity of middle school mathematics and science teachers in the state's Pueblo School District 60. ETS's involvement in this project is twofold: (1) the university has purchased ETS's Keeping Learning on Track® (KLT™) Program and (2) the project is using ETS-developed diagnostic items in mathematics and science that can be used for classroom-based formative assessment purposes. The goal of the project is to examine the impact on teachers and students of supplementing the KLT program with the diagnostic items. In 2010, project directors were planning these activities:

- Collect base-line survey data from students and teachers prior to the KLT training
- Deliver the KLT training to teacher leaders in the district who will then form school-based teacher learning communities
- Provide the supplementary training and diagnostic items in the summer
- Collect ongoing data from teachers to understand programmatic impact of adding the supplementary materials

Measuring the Development of Vocabulary and Word Learning to Support Content Area Reading and Learning

Principal Investigator: Paul Deane

Funding Source: U.S. Department of Education's Institute of Education Sciences

The goal of this project is to develop improved methods for measuring vocabulary and word learning in specific subject areas such as social studies or science. It exploits natural language processing technologies to develop detailed maps of the vocabulary demands of different kinds of texts, and uses these maps to define a sampling strategy for assessing breadth of vocabulary knowledge. There are three main deliverables:

1. Topic maps based upon a statistical analysis of texts in a specific content areas
2. A well-validated battery of item types for assessing partial vocabulary knowledge, making it possible to obtain much more precise measurement when (as is often the case) students have not yet fully mastered more advanced or difficult vocabulary
3. A set of assessments designed to measure breadth and depth of vocabulary knowledge in two specific domains (world history & biological science) at the middle school level

The focus is on subject-matter instruction at the middle school level (the time when domain-specific vocabularies first play an important role in instruction) with the specific goal of developing assessments of vocabulary knowledge in middle school social studies and science.

Work on the project is proceeding in two strands: (1) development of hierarchical topic analyses and (2) development of a set of tasks to measure vocabulary depth). In 2010, project directors were planning these activities:

- Development of hierarchical topic analyses –
 - Complete review of corpus generated by algorithms used to define word associates between topically important words
 - Complete validation of topic hierarchies by content experts
 - Implement software to document the salient topics in the collection of lesson plans and encyclopedia articles covering the major topics of the two domains (world history and biological science)
 - Design the world history and biological science assessments using the data provided by the earlier studies and software
- Development of a set of tasks to measure vocabulary depth
 - Complete analysis of data from two pilot tests (pre- and post-test) administered in Year 1 with an independent, parallel project (Word Generation project lead by Catherine Snow)
 - Take items from the pilot studies and administer them in an in-subject study to evaluate item performance and interaction
 - Analyze the in-subject assessment data

Keeping Learning on Track® Program High School Project

Principal Investigator: Cynthia Tocci

Funding Source: Bill & Melinda Gates Foundation (This work is funded under “Using Assessment to Help Students Succeed” grant, which is directed by Thomas Van Essen)

The goal of the Keeping Learning on Track® Program is to improve the assessment for learning practices of high school math and science teachers and to impact student learning. The KLT program is a sustained, interactive two-year professional development program that supports teachers to adopt assessment for learning strategies that research has shown to greatly increase student learning. Through initial workshops followed by sustained engagement in school-based teacher learning communities, KLT introduces teachers to a wide range of classroom techniques, all unified by the idea: using evidence of student learning to adapt real-time instruction to meet students’ immediate learning needs. As a result of the activities in 2010 ETS hopes to have made an impact in several high need school districts and to have gained additional knowledge about the operational impact of KLT. In 2010, project directors were planning these activities:

- Complete a third administration of the assessment for learning vignettes with the teachers who will have experienced the program for two full years
- Survey the teachers on their program experience including having them estimate the time committed to enact the program

Advanced Psychometrics

An important part of ETS's mission is to advance the field of psychometrics. Much of that work is funded through the ETS research allocation process. However, it is a goal of the Center for External Research to secure outside funding for theoretical and methodological work that is of general interest to the educational research community.

IEA-ETS Research Institute (IERI)

Principal Investigator: Irwin Kirsch / Eugene Gonzalez

Funding Source: International Association for the Evaluation of Educational Achievement (IEA) / Educational Testing Service

The IERI is a collaborative effort between ETS and the IEA Data Processing and Research Center that focuses on improving the science of large-scale assessments. The IERI undertakes activities around three broad areas of work that include research studies related to the development and implementation of large-scale assessments; professional development and training; and dissemination of research findings and information gathered through large-scale assessments. The label "virtual" is attached to the research area to emphasize that the research projects are hosted in the funding institutions, but are facilitated by Web-based collaboration as well as by the shared joint expertise of researchers involved in work on large-scale assessments. The aim of this virtual research area is to contribute to the science of large-scale assessments so that the best available information is provided to policy makers and researchers from around the world. Major activities to be performed in 2010:

- Hold a two-day IERI Large Scale Assessment Conference that will bring together international experts to address *Methodological Issues around Large Scale Assessments*
- Hold spring IERI Academy
- Release a third volume in its monograph series on large scale assessment

At nonprofit ETS, we advance quality and equity in education for people worldwide by creating assessments based on rigorous research. ETS serves individuals, educational institutions and government agencies by providing customized solutions for teacher certification, English-language learning, and elementary, secondary and post-secondary education, as well as conducting education research, analysis and policy studies. Founded in 1947, ETS develops, administers and scores more than 50 million tests annually — including the TOEFL® and TOEIC® tests, the GRE® test and *The Praxis Series*™ assessments — in more than 180 countries, at over 9,000 locations worldwide.

www.ets.org



Listening. Learning. Leading.®

www.ets.org