



# 2011-2012

Research Portfolio  
*ETS Research & Development*

Copyright © 2011 by Educational Testing Service. All rights reserved. ETS, the ETS logo, LISTENING. LEARNING. LEADING., ETS PERSONAL POTENTIAL INDEX, E-RATER, GRE, GRE, TOEFL and TOEIC are registered trademarks of Educational Testing Service (ETS). EXADEP, PRAXIS, and TOEFL IBT are trademarks of ETS. SPEECHRATER and C-RATER are service marks of ETS. All other trademarks are property of their respective owners. 06/08/2011 v.2.0

## Contents

*Editor's note: The contents of this portfolio were current as of June 2011. The status of the projects and initiatives described herein are subject to change at any time. An updated version of this portfolio is expected to be published in early 2012. For more information on the status of research described in this publication, write to us at [RDWeb@ets.org](mailto:RDWeb@ets.org).*

<b>Introduction .....</b>	<b>5</b>
<b>Section I: Research Funded by the ETS Strategic Research Investment .....</b>	<b>6</b>
<i>Cognitively Based Assessment of, for, and as Learning .....</i>	<i>6</i>
<i>Understanding and Enhancing Teaching .....</i>	<i>7</i>
<i>English Language Learning .....</i>	<i>7</i>
<i>Next Generation Higher Education .....</i>	<i>8</i>
<i>Constructed-Response Design and Scoring .....</i>	<i>9</i>
<i>New Constructs .....</i>	<i>10</i>
<i>Validity .....</i>	<i>10</i>
<i>Equating and Applied Psychometrics .....</i>	<i>11</i>
<i>Foundational Statistical and Psychometric Research .....</i>	<i>12</i>
<i>Psychometric Infrastructure.....</i>	<i>13</i>
<b>Section II: Research With Funding From Testing Programs.....</b>	<b>14</b>
Research Studies in 2011 .....	16
<i>Research needed to comply with industry standards for technical quality and/or maintain the ETS brand (with a specific application to a current assessment or product).....</i>	<i>16</i>
<i>Research needed to improve quality, cost-efficiency, or the ETS brand (with a specific application to a current assessment or product) .....</i>	<i>18</i>
<i>Research needed to develop prototypes and/or apply new processes and capabilities to implement an updated or new version of an existing assessment or product.....</i>	<i>18</i>
<i>Research needed to develop prototypes and/or apply new processes and capabilities to launch a new assessment or product .....</i>	<i>19</i>
<i>Research on English language learners (ELLs) in the United States, higher education, and teacher effectiveness.....</i>	<i>20</i>

<b>Section III: Research With Funding From External Agencies .....</b>	<b>21</b>
1. Researching and developing innovative K–12 assessments .....	22
2. Researching and developing a next generation of teacher assessments .....	29
3. Researching and developing improved assessments for English language learners in the United States and internationally .....	33
4. Researching and developing improved assessments for higher education .....	35
5. Extending psychometric methodologies for analysis and reporting of global large-scale group-score assessment results; identifying research questions to be addressed with the results .....	35
6. Researching and developing model assessments that are fair and accessible for students with disabilities .....	39
7. Advancing foundational knowledge, efficiency, and innovation in (a) psychometrics; (b) test development methods; and (c) validity methodology .....	41

## Introduction

This portfolio describes the breadth of the research that the ETS Research & Development division is conducting in 2011. The portfolio includes three main sections, organized according to the three main sources that fund this research:

- (1) The ETS Strategic Research Investment
- (2) Testing programs at ETS
- (3) External governmental and private agencies

For research in each funding area, this portfolio provides information about specific research projects that were active at the time of the portfolio's publication (June 2011), the focus and purpose of the research, and the R&D staff responsible for it. The portfolio also describes how the research aligns with ETS's mission and how it builds organizational knowledge and capability.

Regardless of funding source, ETS R&D research activities 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
- (3) Using R&D knowledge and capabilities to contribute to the development of new assessments

Work in these categories is interrelated: The knowledge and capabilities generated in the first category further help to enhance existing products and also to develop new ones.

## Section I: Research Funded by the ETS Strategic Research Investment

At the time of this portfolio's publication, the ETS Strategic Research Investment was supporting ten initiatives, or programs of research. Below are brief descriptions of the work ETS was undertaking in conjunction with each initiative:

### *Cognitively Based Assessment of, for, and as Learning*

<b>Focus</b>	Researching and developing innovative K-12 assessments
<b>Description</b>	<p>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 build upon cognitive research, state standards, and curricular considerations. Work in 2011 will include:</p> <ul style="list-style-type: none"> <li>• Continued refinement 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)</li> <li>• Creation and field testing of prototype tasks, assessment modules, and school-year assessment designs</li> <li>• Psychometric modeling of task and assessment performances within and across periodic accountability assessments</li> <li>• Development and adaptation of automated scoring models as appropriate for designated task models</li> <li>• Analysis and design of tools to report test information</li> </ul>
<b>Importance</b>	<p>ETS believes that it is important to conduct this research in order to:</p> <ul style="list-style-type: none"> <li>• Solve a pressing educational problem (i.e., creating a balanced assessment system that gathers useful information for policy purposes and effectively supports classroom learning)</li> <li>• Advance the field of educational measurement by developing scientifically sound assessments that teachers consider to be educationally worthwhile</li> </ul>
<b>ETS Contacts</b>	Randy Bennett

### ***Understanding and Enhancing Teaching***

<b>Focus</b>	Researching and developing a next generation of teacher assessments
<b>Description</b>	<p>This initiative focuses on the generation of knowledge and capability to build improved measures of teacher effectiveness. In 2011 the initiative will focus on:</p> <ul style="list-style-type: none"> <li>• Development of assessments of content knowledge for teaching</li> <li>• Analysis of several teacher observation protocols</li> <li>• Scoring designs and scoring data from videos of teachers teaching</li> </ul>
<b>Importance</b>	ETS believes that this research advances its mission as a nonprofit organization by enabling educators to teach more effectively so that their students will succeed in school and in life.
<b>ETS Contact</b>	Courtney Bell

### ***English Language Learning***

<b>Focus</b>	Researching and developing improved assessments for English language learners in the United States and internationally
<b>Description</b>	<p>The purpose of this initiative is 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 2011 research in the initiative will focus on:</p> <p><i>For ELLs in the U.S.:</i></p> <ul style="list-style-type: none"> <li>• Continued research leading to (a) an improved English language proficiency (ELP) assessment; (b) improved subject matter assessments for ELLs; (c) improved teaching of English and other content to ELLs</li> </ul> <p><i>For international ELLs:</i></p> <ul style="list-style-type: none"> <li>• Continued development of a framework for designing an assessment of English proficiency for young learners of English as a foreign language</li> <li>• Investigation of how familiar students of English as a foreign language are with technology and with testing on a computer</li> </ul>
<b>Importance</b>	ETS believes that this research advances its mission as a nonprofit organization by promoting learning and educational performance for all people worldwide.

**Next Generation Higher Education**

<b>Focus</b>	Researching and developing improved assessments for higher education
<b>Description</b>	<p>The purpose of this initiative is to expand access to and improve outcomes in higher education through the conceptualization of and piloting of new assessments and measures. In 2011, research in the initiative will focus on:</p> <ul style="list-style-type: none"><li>• <i>Factors related to access and graduation</i> (projects will explore issues such as: characteristics of students who persist in higher education; skills needed for college graduates to succeed in a global workforce; assessments for adults, including adult immigrant learners, who are preparing to enter or re-enter college)</li><li>• <i>Factors related to outcomes and accountability</i> (projects will explore issues such as: student motivation in taking college outcomes assessments; factors affecting the development of writing skills in college)</li></ul>
<b>Importance</b>	Addressing problems of expanding access, increasing retention and graduation rates, and improving outcomes in higher education are major strategic priorities not only for ETS but also for the United States, as articulated in the recommendations by the Commission on the Future of Higher Education and the Obama Administration's <i>American Graduation Initiative</i> .
<b>ETS Contact</b>	Patrick Kyllonen

### **Constructed-Response Design and Scoring**

**Focus** Researching and developing improved engines for the automated scoring of textual and spoken responses

**Description** The purpose of this initiative is to advance the understanding of fundamental characteristics of how expert human raters perceive, interpret and evaluate responses to CR items; to use this understanding to improve human scoring; and to conceptualize, develop, deploy and improve automated systems that might complement or in some cases replace the work of human evaluators.

A major goal of the automated scoring portion of this effort includes improving the ability of ETS's automated scoring engines to (a) handle *meaning* in a text in a deeper and more flexible way; (b) more fully address the construct of speaking proficiency and produce more reliable and valid scores for existing high-stakes speaking assessments such as TOEFL and TOEIC

With respect to the improved ability to use automated systems to score *text*, work in 2011 will focus on:

- Improvements in the features that the e-rater® and c-rater<sup>SM</sup> scoring engines are able to score
- Investigation of ways to detect organizational structure, factual representation, and opinion in a text

With respect to the improved ability to use automated systems to score spoken responses, work in 2011 will focus on:

- Identification of new features of speech that the SpeechRater<sup>SM</sup> scoring engine is able to score
- Investigation of the performance of a new speech recognition system

**Importance** ETS believes that it is important to conduct this research in order 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
- Reduce the cost and effort of scoring constructed response items

**ETS Contacts** Derrick Higgins and David Williamson

### ***New Constructs***

<b>Focus</b>	Researching and developing assessments of noncognitive factors for educational success that have not traditionally been the focus of ETS assessments
<b>Description</b>	<p>The purpose of this initiative is 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. The initiative in 2011 will focus on:</p> <ul style="list-style-type: none"> <li>• Improving the way background factors are assessed in questionnaires on international survey assessments</li> <li>• Identifying applications of the theory of planned behavior</li> </ul>
<b>Importance</b>	<p>ETS believes that it is important to conduct this research in order to:</p> <ul style="list-style-type: none"> <li>• Expand our ability to assess, in valid and reliable ways, constructs other than cognitive ability and academic achievement</li> <li>• Investigate ways in which noncognitive skills levels can be improved</li> </ul>
<b>ETS Contact</b>	Richard Roberts

### ***Validity***

<b>Focus</b>	Ensuring technical quality for existing and new assessments for all individuals
<b>Description</b>	<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"> <li>• Support the psychometric quality of new and established ETS tests and products</li> <li>• Establish validity, fairness, and accessibility of assessments for students with disabilities</li> <li>• Expand standard-setting and job analysis methodology</li> </ul>
<b>Importance</b>	<p>ETS believes it is important to conduct this research in order to:</p> <ul style="list-style-type: none"> <li>• Address the ETS mission's call to create assessments and assessment-related products that are fair for all learners</li> <li>• Respond to greater public demand for scientific evidence of the efficacy of ETS's assessment products and services</li> </ul>
<b>ETS Contact</b>	Brent Bridgeman

### ***Equating and Applied Psychometrics***

<b>Focus</b>	Improved methodology in the areas of equating and applied psychometrics for the benefit of ETS programs and the profession in general
<b>Description</b>	<p>The purpose of this initiative is 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, the initiative also includes research on other applied methodologies. In 2011, the initiative will:</p> <ul style="list-style-type: none"><li>• Identify methodologies to ensure the stability and meaning of reported scores over time</li><li>• Improve current equating and linking methodology (comparing equating methods, evaluating smoothing methods, and studying effects of sampling on equating)</li><li>• Improve analysis related to subscore reporting</li><li>• Improve communication of analyses results to clients</li></ul>
<b>Importance</b>	<p>ETS believes it is important to conduct this research in order to:</p> <ul style="list-style-type: none"><li>• Ensure the quality, equity, and fairness of assessments</li><li>• Enhance the efficiency of testing programs</li><li>• Help clients better understand how analyses results should be interpreted</li></ul>
<b>ETS Contacts</b>	Alina von Davier and James Carlson

***Foundational Statistical and Psychometric Research***

<b>Focus</b>	Fundamental contributions to ETS and to the profession in the areas of statistics and psychometrics
<b>Description</b>	<p>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:</p> <ul style="list-style-type: none"><li>• Applications of item-response theory to complex problems; development of new modeling approaches for mixed response data from computerized assessments</li><li>• The methodology to examine the agreement of models to data</li><li>• Methods for aggregation of examinee data from multiple assessments</li><li>• The assessment of variability of estimates under complex sampling</li><li>• Methods to avoid excessively complex models</li></ul>
<b>Importance</b>	<p>ETS believes that it is important to conduct this research in order to:</p> <ul style="list-style-type: none"><li>• Ensure that the methodology used in psychometric operations is defensible and efficient in both a computational and statistical sense</li><li>• Advance the field of educational measurement</li></ul>
<b>ETS Contact</b>	Shelby Haberman and Matthias von Davier

### ***Psychometric Infrastructure***

<b>Focus</b>	Properly and efficiently supporting current and new testing programs through practical development of software and data processing methods
<b>Description</b>	<p>This initiative focuses on the development of statistical/psychometric infrastructure to increase operational and computational efficiency, and prevent errors. In 2011, projects in the initiative will:</p> <ul style="list-style-type: none"><li>• Standardize psychometric processes across programs and work groups</li><li>• Continue enhancement of NAEP operational software, including both the operational systems used at ETS and DESI software</li><li>• Add to or improve our general data processing hardware and software capabilities</li></ul>
<b>Importance</b>	<p>ETS believes that it is important to conduct such research in order to:</p> <ul style="list-style-type: none"><li>• Increase confidence in the integrity and repeatability of the results produced from ETS’s statistical systems</li><li>• Improve operational and computational efficiency of ETS’s software and data processing methods</li></ul>
<b>ETS Contact</b>	Tim Davey

## Section II:

### Research With Funding From Testing Programs

ETS Research & Development conducts research studies and performs other activities that ETS testing programs fund. R&D carries out this work in addition to the research funded by the internal research allocation. Specific research and other activities are defined for each program in order to:

- Help ensure the technical and psychometric quality of the assessments in the program
- Aid in meeting the internal ETS audit standards
- Assist in the prioritization of needed research on a program level
- Determine the fundamental work needed for revising a program

Listed below are some of the testing programs for which R&D was conducting or planning to conduct research at the time of this portfolio's publication in June 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 program's research concerns. 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 their public release that they are completed on time, within budget, and according to technical standards.

Program	Research Liaison
College Board Programs	John Young
GRE® General and Subject Tests	Brent Bridgeman
High Schools That Work (HSTW)	John Young
K-12 Programs	Cara Laitusis
ETS Proficiency Profile	Lydia Liu
Major Field Tests (MFT)	Guangming Ling
PRAXIS™	Richard Tannenbaum
Texas	Richard Tannenbaum
TOEFL®	Xiaoming Xi and Donald Powers
TOEFL® Junior™	Mikyung Wolf
TOEFL® Primary Development	Sultan Turkan
TOEIC®	Donald Powers
EXADEP	Brent Bridgeman

In addition, R&D 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 psychometric manager ensures that each program maintains the highest possible psychometric quality.

These are the psychometric managers for 2011:

#### College Board Programs

- *AP*®, *CLEP*, *SAT Subject Tests*  
Psychometric Manager: Ourania Rotou
- *SAT*, *PSAT*®, *NMSQT*®  
Psychometric Manager: Jinghua Liu

#### Higher Ed Programs

- *GRE General Test*  
Psychometric Manager: Fred Robin
- *Texas*  
Psychometric Manager: Gautam Puhan
- *PRAXIS*  
Psychometric Manager: Kevin Larkin
- *ETS Proficiency Profile*, *MFT*,  
*GRE Subjects Tests*  
Psychometric Manager: Michael Walker

#### K-12 Programs

- *HSTW*, *MGA*, *CBAL*, *ERB*, *Tennessee*,  
*Maryland*, *Chicago*, *CSU*, *EAP*, *Texas*  
Psychometric Manager: Christine Mills
- *CAHSEE*, *WCAP*, *CPS*, *SREB*  
Psychometric Manager: Hyeonjoo Oh
- *CAPA*, *CST*, *CMA*, *STS*  
Psychometric Manager: Lixiong Gu

#### Global and Workforce

- *TOEFL*, *TOEFL Junior*  
Psychometric Manager: Terran Brown
- *EXADEP*, *TOEIC*  
Psychometric Manager: 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.

## Research Studies in 2011

In 2011, ETS testing programs will support research that falls into four major categories. These high-level categories describe the purpose of the research that R&D carries out and the impact and role that this research plays in supporting the ETS brand. The four major categories of research are:

- (a) Research needed to comply with industry standards for technical quality and/or to maintain the ETS brand (with a specific application to a current assessment or product)
- (b) Research needed to improve quality, cost-efficiency, or the ETS brand (with a specific application to a current assessment or product)
- (c) Research needed to develop prototypes and/or apply new processes and capabilities to implement an updated or new version of an existing assessment or product
- (d) Research needed to develop prototypes and/or apply new processes and capabilities to launch a new assessment or product

### ***Research needed to comply with industry standards for technical quality and/or maintain the ETS brand (with a specific application to a current assessment or product)***

This category of research includes foundational and applied psychometric research to maintain validity and fairness for ETS assessments. It also includes research to ensure appropriate technical quality for existing ETS assessments for all individuals. In addition, work on standard setting practices and improvements fall in this category.

Some of this research provides validity evidence by documenting support for 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 of critical importance. It is also critically important to ensure that the test is fair for all subpopulations of examinees, including groups defined by nationality, race/ethnicity, gender, disability status, and English language proficiency levels.

Research to ensure the technical quality of ETS assessments includes statistical and psychometric work on the soundness and meaning of scores.

Some of the validity work that ETS R&D staff members are conducting in 2011 involves efforts to build validity rationales for testing programs, such as:

- An evaluation of the predictive validity of the ETS Personal Potential Index to support its use in higher education admissions decisions
- Research on the validity of the GRE in predicting success in graduate school, using state-level data
- A study investigating the impact of domestic ELL status on outcomes assessment scores for the ETS Proficiency Profile

- Validity studies for the High Schools That Work (HSTW) assessment

Other work will involve research to ensure fairness in assessments for various testing programs, including:

- A study to establish best practices for demonstrating fairness for the TOEFL test
- Examining whether the revised GRE predicts graduate school success just as well for students with disabilities relative to the general population

Research in this category also aims to maintain the technical quality of ETS assessments. An example would be:

- An investigation of score comparability across essay prompts for the College Level Examination Program (CLEP)

Technical quality is also maintained by conducting scaling and linking studies to ensure that scores retain their meaning over time as test taker populations and the assessments change. Examples include:

- Linking international exam forms for the Advanced Placement (AP) program
- An investigation of the invariance of linking functions for the SAT Reasoning Test (SAT)
- An investigation of the comparability of TOEFL integrated writing prompts
- Work to provide a one-time-use concordance table for the purpose of connecting the old scale to the new scale for several assessments from the Washington Comprehensive Assessment Program (WCAP)

In addition to the studies highlighted above, R&D staff will, in 2011:

- Conduct several studies on score meaning and use internationally
- Determine how best to report student growth for California's accountability assessment
- Explore the use of GRE scores in awarding scholarships to graduate students, in order to build a foundation for a future validity study
- Conduct standard setting studies to create cutscores that are valid and interpretable. Specifically, researchers will determine appropriate cutscores for the WCAP math and science assessments and continue PRAXIS work using the newly-created regional standard setting methodology and virtual standard setting approach

***Research needed to improve quality, cost-efficiency, or the ETS brand (with a specific application to a current assessment or product)***

This category includes research to improve current methodologies and practices for scoring and equating, research to improve the efficiency by which assessments are scored, and work to expand the use of ETS assessments.

Research studies to improve current methodologies include:

- A study to determine whether an alternate method of equating TOEIC test forms can be used
- Evaluation and development of methods to provide diagnostic feedback on TOEIC
- Research on effective ways to rescale SAT subject tests
- An evaluation of rights scoring for Advanced Placement (AP) tests

R&D will also study the feasibility of changing modes of delivering assessments. Examples include:

- A study on the comparability of computer-based and paper-based Praxis I® tests
- Research on the comparability of scores from paper-based and computer-based TOEIC tests delivered in China
- An evaluation of different administration modes for WCAP

***Research needed to develop prototypes and/or apply new processes and capabilities to implement an updated or new version of an existing assessment or product***

This category of work includes research to ensure the technical quality of updated, existing ETS assessments. Such research includes evaluating the efficacy of scaling procedures for the revised GRE test. Linking studies will also be carried out for different types of exams to allow users to compare scores from various assessments. This work supports the wider use of ETS assessments and includes the creation of a tool to link the revised GRE to GMAT scores.

In addition, several research studies will be conducted in 2011 to help expand the ETS market. Studies include:

- Determining what gives an English assessment global appeal
- Exploring, in cooperation with the University of Maastricht, the use of the GRE tests in Europe for higher education admissions
- An investigation of the feasibility of using the Major Field Tests (MFT) in business programs abroad

***Research needed to develop prototypes and/or apply new processes and capabilities to launch a new assessment or product***

This category of work includes research to ensure the technical quality of new ETS assessments. Research in 2011 will focus on a number of new assessments. Examples of this work include:

- Foundational research for TOEFL Junior, including work on score meaning and studies on content, construct, and predictive validity
- Evaluating of the impact of the Engaging English product on TOEFL Junior reading performance
- Developing assessment design and analysis models for the TEL NAEP assessment

In addition to research support for new ETS assessments, R&D staff will conduct studies in 2011 in preparation for future assessments that are being discussed in the education community. Such work includes:

- Research supporting Next Generation Teacher Assessments, including the development of new item types for knowledge for teaching and Tier II Praxis licensure assessments
- Determining state models for evaluating teachers, principals, and superintendents in the state of Kansas
- Developing automated scoring models, both speech and text, for the English as a Foreign Language (EFL) Teacher Certification test

In addition, researchers will provide support for K–12 for the development of proposals related to the common core standards efforts.

### ***Research on English language learners (ELLs) in the United States, higher education, and teacher effectiveness***

In addition to the four major categories of research described above, several research studies that R&D staff will conduct in 2011 fall into one of the three key areas identified by the Board of Trustees. The following studies can be classified in the areas of domestic ELLs, higher education, and teacher effectiveness.

#### **ELLs in the United States**

- A study investigating the impact of ELL status on outcomes assessment scores for the ETS Proficiency Profile

#### **Higher Education**

- Validity analyses for the ETS Personal Potential Index
- Validity analyses for the revised GRE for students with disabilities
- Validity analyses using GRE scores from a state-level database
- Research on expanding the MFT in Business to international business programs
- A study on linking scores from the revised GRE to the GMAT
- Research to explore the use of the GRE in Europe higher education admissions in cooperation with the University of Maastricht

#### **Teacher Effectiveness**

- Research supporting Next Generation Teacher Assessments

## Section III: Research With Funding From External Agencies

The Center for External Research in the ETS Research & Development Division coordinates more than 35 externally funded research projects<sup>1</sup> directed in some way by ETS scientists. Funding agencies include various divisions of the U.S. Department of Education (for example, the Institute of Education Sciences and the Office of Special Education Programs), the National Institute of Child Health & Human Development, the Organisation for Economic Co-operation & Development, Westat, and The Bill & Melinda Gates Foundation.

In determining what externally funded research to undertake, 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, the division is cognizant that its external research activities can advance other enterprises at ETS.

Externally funded research allows ETS scientists to develop relationships with other practitioners, expand their own expertise, and augment the organization's mission-related research.

The work below organizes the division's external research activities according to these categories:

- (1) Researching and developing innovative K–12 assessments
- (2) Researching and developing a next generation of teacher assessments
- (3) Researching and developing improved assessments for English language learners in the United States and internationally
- (4) Researching and developing improved assessments for higher education
- (5) Extending psychometric methodologies for analysis and reporting of global large-scale group-score assessment results; identifying research questions to be addressed with the results
- (6) Researching and developing model assessments that are fair and accessible for students with disabilities
- (7) Advancing foundational knowledge, efficiency, and innovation in (a) psychometrics; (b) test development methods; and (c) validity methodology

In the sections that follow, we describe the goals of the current projects within these areas and work projected for 2011. The research of some projects fits across multiple categories. In these instances the full description is in the category that most fits the aim of the project. For further information on the projects listed, please contact Thomas Van Essen, Executive Director of the Center for External Research ([tvanessen@ets.org](mailto:tvanessen@ets.org)).

---

<sup>1</sup> Many of the smaller projects are not featured in this document.

## 1. Researching and developing innovative K–12 assessments

### *Developing Assessments for the Reading for Understanding Initiative*

**Principal Investigators:** John Sabatini / Tenaha O'Reilly

**Funding Source:** U.S. Department of Education Institute of Education Sciences (IES)

ETS is leading one of the six research teams participating in the Institute of Education Sciences' Reading for Understanding Research Network. Together, more than 130 researchers are working to improve reading comprehension for all students from preschool through high school. The project is part of a five-year IES effort to advance the science of reading instruction with the goal of enabling all students to fully comprehend written material. ETS is leading the team that is developing the next generation of pre-K through 12th grade reading assessments. On the project, ETS is working with Florida State University, Northern Illinois University, and Arizona State University. The ETS team is developing reading assessments that are aligned with contemporary theoretical constructs and sensitive to changes in skill development. The assessments also will emphasize strategic reading processes and provide greater information for guiding instruction. The other network members are focusing on instructional strategies, professional development, and technology. For more information on the Reading for Understanding Research Initiative, visit <http://ies.ed.gov/ncer/projects/program.asp?ProgID=62>.

In 2011 the project directors are planning the following activities:

- Continue to design, program, and administer prototype assessments
- Integrate findings from performance-level descriptor study with assessment design
- Integrate focus group findings with assessment design
- Modify existing designs based on validity and construct studies
- Prepare draft of a model and theory of how reading develops over time

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

**Principal Investigators:** John Sabatini / Jane Shore

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

The central goal of this project is to estimate the prevalence of various types of reading difficulties 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. 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 difficulty and for each subtype of reading difficulty. It is anticipated that the reading difficulty profiles will include deficits, alone and in combination, in comprehending text, recognizing/decoding words, and reading fluency. Each reading difficulty 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.

In 2011 the project directors are planning to

- Recruit 25 elementary schools and 37 middle schools for the effort to have some 1,630 fourth graders and 2,000 eighth graders participate in the study
- Distribute parent questionnaires

### ***Study Aid and Reading Assistant (SARA)***

**Principal Investigators:** John Sabatini / Kelly Bruce

**Funding Source:** Multiple sources

SARA is a suite of diagnostic assessments based on a componential understanding of reading. SARA is built on a research base of studies from the fields of cognition, linguistics, and neuroscience. The assessment suite 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 the Strategic Education Research Partnership (SERP, Inc.) being the most notable. SARA also is in use in the Programme for the International Assessment for Adult Competencies (PIAAC).

In 2011 project directors are planning to complete at least two large-scale data collection efforts and several small-scale, pilot data collections. The large-scale data collections will take place in (1) a large urban district in the mid-Atlantic region of the United States, with an expected 10,000 students across three grades participating; and (2) a medium-size urban school district in the northeastern United States with an expected 4,000 students across four grades.

### ***Creating Cross-Grade Assessments of the Development of Core Algebraic Constructs***

**Principal Investigators:** Malcolm Bauer / Caroline Wylie

**Funding Source:** U.S. Department of Education Institute of Education Sciences

With this four-year project, ETS is bridging the gap between educational research and classroom practice by developing a set of innovative algebra assessments that middle school teachers can use to improve student learning and instruction.

Research is beginning to suggest that teachers are struggling to determine which instructional steps to take based on the formative assessment evidence they gather from their students. The aim of the project is to provide rich profiles of groups of students with respect to the models of mathematical development to support teachers' day-to-day instructional decision making. The set of formative assessments being developed is focusing on three key algebraic concepts — equality, notion of a variable, and multiplicative reasoning — and is being designed to be readily adaptable to fit in with teachers' current curricula. The research plan calls for the assessments to be field tested with a national sample of 4,000 to 6,000 students in grades 6 through 8. The project is expected to be completed by June 2014.

In 2011 project directors are planning to complete the development of initial item pool for developmental models in preparation for the second phase of the effort.

### ***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 it uses these maps to define a sampling strategy for assessing breadth of vocabulary knowledge. There are three main deliverables:

- (1) Topic maps based on a statistical analysis of texts in 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 and 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.

Major work to occur in 2011 includes:

- Performing data analysis to determine the robustness of the item types
- Administering pilot tests focused on domain vocabulary in middle school biology and history

### ***Assessing Noncognitive Characteristics in Middle School Students***

**Principal Investigators:** Jeremy Burrus / Richard Roberts

**Funding Source:** Benchmark Research (Elementary Schools Research Collaborative)

ETS is conducting a three-year longitudinal study to examine how effective schools in the Elementary Schools Research Collaborative (ESRC) are in improving their students' noncognitive skills. ETS is creating or adapting existing measures of noncognitive characteristics for the effort. Some 20 schools in the ESRC are participating. Each has agreed to administer the assessment to at least 100 middle school students twice a year over the course of the study. The skill set measured by the assessments will include noncognitive characteristics such as resilience, ethics, integrity, creativity, time management, and teamwork — all viewed as important to achieving success in academics and in the workplace. In 2011 the project directors plan to:

- Meet with representatives of the ESRC to make decisions concerning the constructs being measured and the methods for measuring them
- Work with the ESRC to certify compliance with guidelines to protect human subjects involved in the research
- Develop a set of assessments to measure the preselected characteristics for middle school students
- Conduct cognitive labs prior to the administration of the assessments to assess student understanding of the items and adjust any items viewed as being problematic
- Code the items for Web-based administration
- Have the participating schools conduct the first administration of the assessments in November 2011

### ***Cumulative Learning Using Embedded Assessment Results (CLEAR)***

**Principal Investigator:** Lydia Liu  
**Funding Source:** National Science Foundation  
(ETS is subrecipient to the University of California, Berkeley)

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 (STEM) course more efficient.

The project team is designing pretests, posttests, and 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.

Major activities to be performed in 2011:

- Design annual posttest 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 posttest 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:** National Science Foundation  
(ETS is subrecipient to the University of California, Berkeley)

This project is a study of the effectiveness of interactive, online animations or visualizations — which include models, simulations, and virtual experiments — in affecting instruction by making unseen scientific phenomena, such as chemical reactions, visible to learners. VISUAL aims to transform instruction in physical science by:

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

The goals of the project in 2011 are to

- Develop an assessment that will provide a baseline measure of students' science proficiency and that 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

### ***Early Childhood Longitudinal Study (ECLS-K:11)***

**Principal Investigator:** Michelle Najarian  
**Funding Source:** U.S. Department of Education Institute of Education Sciences  
(ETS is a subcontractor to Westat)

ETS is a subcontractor to Westat on this, the third in a series of longitudinal studies sponsored by the U.S. Department of Education 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 is designed to allow cross-cohort comparisons of two nationally representative kindergarten classes experiencing different policy, educational, and demographic environments. The study is following a nationally

representative cohort of children who started kindergarten in the 2010–2011 school year. There are two phases to this work. For Phase I, in 2011 the project director is planning to:

- Complete analysis of test data from the fall kindergarten national assessment in reading and mathematics
- Complete analysis of the test data from the spring kindergarten national assessment in reading, mathematics, and science
- Begin analysis of test data from the fall first grade national assessment in reading, mathematics, and science

For Phase II, the project director is planning to finalize test forms in reading, mathematics, and science for first grade for the national data collections in 2011–2012.

### ***National Title I Study of Implementation and Outcomes: Early Childhood Language Development***

**Principal Investigator:** Michelle Najarian

**Funding Source:** U.S. Department of Education Institute of Education Sciences  
(ETS is a subcontractor to Mathematica Policy Research)

The effort is a national study that involves preschool and early elementary programs (K–3) in Title I schools. This project is intended to collect data on programs and classroom practices as well as child outcomes for the purpose of identifying programs and classroom practices associated with children’s language development, background knowledge, and comprehension outcomes. ETS’s role is to study the psychometric properties of reading items from the Early Childhood Longitudinal Study (ECLS) program for their inclusion in the ECLS-K:11 second grade field test and the ECLS-K:11 third grade national assessment. ETS also will be involved in the scoring and data analysis of the assessments. In 2011 major activities include:

- Selecting passages and associated item sets from the ECLS-K:11 first and second grade forms in reading to administer to the sample;
- Beginning analysis of the test data from the fall 2011 national assessments in reading and general knowledge.

### ***Head Start Family and Child Experiences Survey (FACES)***

**Principal Investigator:** Michelle Najarian  
**Funding Source:** U.S. Department of Health & Human Services  
(ETS is a subcontractor to Mathematica Policy Research)

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 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 2011 the project director is planning to analyze test data from the spring 2011 national assessments in math and letter recognition.

## **2. Researching and developing a next generation of teacher assessments**

### ***Understanding Teacher Quality***

**Principal Investigators:** Drew Gitomer / Courtney Bell  
**Funding Source:** The Bill & Melinda Gates Foundation

ETS is leading a collaboration that is studying teaching evaluation systems to gain a fuller understanding of what makes effective practice. The study also is investigating 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/](http://www.metproject.org/)).

The 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.

Major activities to be performed in 2011:

- Complete data analysis
- Analyze the results
- Begin the report writing and dissemination processes

### ***Measures of Effective Teaching / Teacher Knowledge***

**Principal Investigators:** Drew Gitomer / Geoffrey Phelps

**Funding Source:** The Bill & Melinda Gates Foundation

Researchers from ETS and the University of Michigan are designing and developing assessments that measure teachers' knowledge of the specific subject matters they teach. The Bill & Melinda Gates Foundation is supporting the effort through its Measures of Effective Teaching (MET) venture ([www.metproject.org](http://www.metproject.org)). The new measures are based on frameworks that tap into the specific subject-matter knowledge teachers regularly use in their day-to-day practice. Through MET, the researchers have developed frameworks for the assessment of teacher knowledge with the specific focus on content pedagogy and content knowledge. The project team is using the frameworks to develop assessments of teacher knowledge in mathematics and English language arts in grades 4 to 9 that address both general knowledge and specific focal topics. The goal is to have a set of assessments by December 2011. Major activities to be performed in 2011:

- Complete the administration of all test forms
- Conduct an analysis of the pilot test and study results

### ***Toward an Understanding of Classroom Context (CLASS-S Validation Study)***

**Principal Investigators:** Drew Gitomer / Courtney Bell

**Funding Source:** William T. Grant and Spencer Foundations

ETS, the University of Virginia, and The RAND Corporation received a 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 is testing 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 has gathered 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 has collected 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 with the changes in student scores on a standardized algebra test.

Major activities to be performed in 2011:

- Complete analyses
- Write final report and associated journal articles

### ***Video Plus Scoring***

**Principal Investigator:** Catherine McClellan

**Funding Source:** Teachscape (Through a grant from The Bill & Melinda Gates Foundation)

ETS is working with Teachscape to score the videos captured as part of The Bill & Melinda Gates Foundation Measures of Effective Teaching (MET) project ([www.metproject.org](http://www.metproject.org)). ETS is adapting a number of well-known, research-based teaching assessment frameworks to evaluate and score the thousands of videos of teachers’ lessons captured through the Teachscape system. These frameworks include Classroom Assessment Scoring (CLASS), developed by Robert Pianta, University of Virginia; Framework for Teaching, developed by Charlotte Danielson; Mathematical Quality of Instruction (MQI), developed by Heather Hill, Harvard University, and Deborah Loewenberg Ball, University of Michigan; Protocol for Language Arts Teaching Observations (PLATO), developed by Pam Grossman, Stanford University; and Quality Science Teaching (QST) Instrument, developed by Raymond Pecheone, Stanford University. Major activities to be performed in 2011:

- Complete scoring study designs for the five classroom assessment instruments
- Complete video scoring and provide data to Teachscape

### ***Creation & Dissemination of Upper Elementary Math Assessment Modules***

**Principal Investigator:** Judy Hickman  
**Funding Source:** Harvard Graduate School of Education  
(Through a grant from the National Science Foundation)

This 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 students' 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 provide the National Science Foundation's Math-Science Partnerships and others with 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.

Major activities to be performed in 2011:

- Conduct field tests of four to six new Geometry/Numbers Operations test forms with 500 students to get preliminary stats on items
- Administer second operational test (posttest to the fall 2010 pretest) of original four Algebra/Numbers Operations test forms with 6,000 students and 300 teachers
- Conduct growth analysis on the data from the Algebra/Numbers Operations pre- and posttests
- Design 8 to 12 new test forms with vertical linking design for administration in the fall of 2011 and in early 2012

### ***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 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 the National Comprehensive Center for Teacher Quality is to ensure a highly qualified teacher in every classroom. In addition to ETS and LPA, 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](http://www.tqsource.org) provides interactive databases, comprehensive research studies, publications, and other resources. Major activities anticipated in 2011:

- Continue to offer technical assistance to states
- Work with partners to develop long-range plans for the center

### **3. Researching and developing improved assessments for English language learners in the United States and internationally**

#### ***Developing a Formative Assessment of Academic Reading to Improve Learning and Teaching of ELL Students***

**Principal Investigators:** Mikyung Wolf / Jane Shore

**Funding Source:** U.S. Department of Education Institute of Education Sciences

Through this four-year project, ETS is developing a formative assessment of academic reading comprehension for English language learners. ELL students deal with the dual challenges of acquiring English language proficiency to handle academic materials and learning curriculum content. At the higher grade levels, the increased complexity of course materials presents an increasingly difficult learning context. This project is developing and validating a classroom-based, formative assessment of academic reading for ELL students in middle school. The resulting formative assessment tool will consist of a set of assessment forms specifically designed for use with ELLs. Use of the assessments will be supported by a Web-based score reporting tool and guidelines to support the use of the assessments in instruction. Working with ETS on the project is the National Center for Research on Evaluation, Standards, and Student Testing at the University of California, Los Angeles. In 2011 the project directors plan to:

- Develop prototype items/tasks
- Develop pilot forms;
- Conduct pilot and field testing
- Refine the construct and test-specification documents

### ***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 Institute of Education Sciences

The goal of this project is to develop a technology-rich teacher professional development (TPD) package 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 and (2) Instructional Authoring. 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 (renamed Language Muse for this project), 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 2011 the project director is planning these activities:

- Begin to pilot the intervention with the Language Muse system with one group of teachers at George Washington University and with as many as six groups at Stanford University
- Pursue continued development of Language Muse based on teacher use and feedback
- Identify teachers who will let the project follow them into their classrooms to see how the teachers use the Language Muse for materials development

#### 4. Researching and developing improved assessments for higher education

##### *Feasibility Study for an Assessment of Higher Education Learning Outcomes (AHELO) – Economics*

**Principal Investigators:** 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 six 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.

In 2011 project directors are planning these activities:

- Complete framework
- Deliver assessments for feasibility study

#### 5. Extending psychometric methodologies for analysis and reporting of global large-scale group-score assessment results; identifying research questions to be addressed with the results

##### *Programme for the International Assessment for Adult Competencies (PIAAC)*

**Principal Investigators:** Irwin Kirsch / Claudia Tamassia  
**Funding Source:** Organisation 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. The study is measuring 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 is designed to be a measure of the literacy and numeracy skills of people between the ages of 16 and 65. It also looks 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:

- a. 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
- b. Assess the relationship between these competencies and various social, educational, and economic outcomes
- c. Gauge how successful education and training systems are at generating these outcomes
- d. 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 2011 the project directors plan to:

- Finalize the analysis of the field test results
- Provide support to countries that are finalizing their materials for the main study, including the background questionnaire, the cognitive paper- and computer-based items, the orientation modules, and a working Virtual Machine. This will also include providing directions and approval for further revisions and verifying changes
- Deliver national versions of the computer- and paper-based assessments so that the main study can begin in August 2011 and end in March 2012
- Support countries in testing their interview platform for the main study (i.e., Main Study Virtual Machines)
- Organize international training for the main study in June 2011 that will include updated interviewer training, general guidelines for scoring of the paper booklets, and additional information for coding industry and occupation
- Provide support to countries during the implementation of the main study
- Start drafting the PIAAC Technical Report

### ***Programme for International Student Assessment (PISA)***

**Principal Investigators:** Patrick Kyllonen / Irwin Kirsch  
**Funding Source:** Organisation for Economic Co-operation and Development (OECD)  
 (ETS is a subrecipient to the Australian Council for Educational Research)

The Programme for International Student Assessment is a project of the 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:

- a. How well are young adults prepared to meet the challenges of the future?
- b. Are they able to analyze, reason, and communicate their ideas effectively?
- c. Do they have the capacity to continue learning throughout life?
- d. 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, the results of which are comparable across different national and cultural contexts. The survey is implemented in intervals of three years. 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 2011:

- Analyze the field trial data from 70 countries (There were 1,000 participants per country and the background form had more than 500 contextual questions in various traditional and innovative new formats measuring socioeconomic status, student attitudes, classroom practices, school characteristics, parent beliefs and behaviors, and other factors.)
- Prepare reports on the field trial data analysis
- Based on the field trial analysis findings of predictiveness of student achievement and cross-country comparability, decide which items and formats to use for the PISA Main Survey, which will be conducted in 2012.

### ***U.S. Participation in the Programme for International Assessment of Adult Competencies***

**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 2011 the project director is planning these activities:

- Provide initial thoughts on the report outline
- Begin preparations for handling the analysis of the data and drafting the national report

### ***Trends in International Math and Science Study (TIMSS)***

**Principal Investigators:** John Barone / Ted Blew  
**Funding Source:** U.S. Department of Education Institute of Education Sciences  
(ETS is a subcontractor to Boston College)

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.

In 2011 the project directors plan to continue to offer technical support on the maintenance and use of the data set.

### ***IEA-ETS Research Institute (IERI)***

**Principal Investigators:** 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, which include research studies related to the development and implementation of large-scale assessments (research area); professional development and training (training area); and dissemination of research findings and information gathered through large-scale assessments (dissemination area). 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 policymakers and researchers from around the world. Major activities to be performed in 2011:

- 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 and fall IERI Academy
- Release a fourth volume in the monograph series on large-scale assessment

## 6. Researching and developing model assessments that are fair and accessible for students with disabilities

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

**Principal Investigators:** Cara Laitusis / Yigal Attali  
**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. In the process, one distracter (wrong answer choice) is removed after the student attempts the test item, and the student receives 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 who are taking state assessments.

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 few 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.

Awarded in January 2010, the project began with a series of small studies exploring a variety of response and scoring formats, item types, and content areas using cognitive interview methodology. Using the results of these studies, the project team has started preparations to 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. The researchers are taking a broad view of validity to include the psychometric comparability and accuracy of scores. In addition, the project team is examining the perceptions of students and their teachers on the testing experience and usefulness of test scores for instructional decision making. In 2011 the project directors are planning these activities:

- Work with the NCEO in an iterative process, using the results of three NCEO-conducted qualitative studies, to refine the type of feedback that will be most useful to students with disabilities during large-scale assessments
- Continue to revise the computer-based testing platform, test items, and types of feedback
- Begin preliminary planning for a large data collection in 2012

### ***Technology Assisted Reading Assessment (TARA)***

**Principal Investigator:** Cara Laitusis  
**Funding Source:** U.S. Department of Education

The U.S. Department of Education began funding the Technology Assisted Reading Assessment 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) colead TARA.

TARA's program of research includes psychometric research that 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.

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](http://www.naraptara.info), as well as the NARAP website, [www.NARAP.info](http://www.NARAP.info).

In 2011 the project director is planning the following activities:

- Complete the statistical analyses of the TARA field trial
- Incorporate results into a technical report for the project and present to the project Technical Advisory Committee

### ***Development of Computer-Based Testing Accommodations for Students with Visual Disabilities***

**Principal Investigator:** Cara Laitusis

**Funding Source:** U.S. Department of Education Institute of Education Sciences

ETS is working with Measured Progress's Nimble Innovation Lab on a project that will provide a fully functional model of a computer-based test delivery platform explicitly designed to meet the needs of students who are blind or have low vision.

None of the accessible, computer-delivered testing platforms available today allows for use of a braille keyboard or utilizes refreshable braille display. The project will develop testing accommodations that include key-stroke navigation, braille keyboard input of responses, refreshable braille display, screen magnification, and audio output. The focus of the effort will be to deliver an eighth-grade reading assessment with the set of accommodations. However, the true aim of the three-year project is to provide a blueprint for the field on how to design more fully accessible computer-based test delivery platforms.

In 2011 the project director is planning the following activities:

- Develop code and software to upgrade the existing computer-based test delivery platform to enable it to function with added accommodations
- Conduct a series of small pilot tests with students with visual impairments to check the functionality of the modifications

## **7. Advancing foundational knowledge, efficiency, and innovation in (a) psychometrics; (b) test development methods; and (c) validity methodology**

### ***Psychometric Models for 21<sup>st</sup> Century Educational Survey Assessments***

**Principal Investigators:** Frank Rijmen / Matthias Von Davier

**Funding Source:** U.S. Department of Education Institute of Education Sciences

This project is developing psychometric models that are better aligned to the innovative item types that are increasingly in use in large-scale educational survey assessments, such as the National Assessment of Educational Progress (NAEP) and the Programme for International Student Achievement (PISA).

These survey programs and others have evolved over time to better incorporate changes in how students learn — for example, how the use of technology has changed instruction. To keep pace, this three-year project is updating the models used to analyze the data from such survey programs.

Technological advances have brought about the development of innovative item types, including technology-based tasks and integrated tasks that have the potential to tap multiple skills. The same technology also has affected how students learn, so it is to be expected that such item types will become even more prevalent in these programs. This project is focusing on the psychometric models that mirror how current assessment frameworks characterize the relationship between tasks and the content domain being assessed. In addition, the project team is developing models that will be attentive to the complex sampling design used in these studies.

On the project, ETS is working with researchers at the University of California, Berkeley, University of Maryland, Wake Forest University, and the International Association for the Evaluation of Educational Achievement.

Major activities to be performed in 2011:

- Formulation of multidimensional item response theory models that take into account the assessment frameworks
- Determination of computational complexity for maximum likelihood estimation for all formulated item response theory models
- Construction of efficient maximum likelihood estimation methods when possible
- Selection of models applied to NAEP or other large-scale survey assessment
- Dissemination of results of the above activities