Application for Federal Education Assistance (ED 424)

U.S. Department of Education

Form Approved
OMB No. 1875-0106
Exp. 11/30/2004

Applicant Information

Organizational Unit

1. Name and Address

Legal Name:__ Educational Testing Service________________________________

Address: ___ Center for Learning & Teaching Research________________

____ Rosedale Road, MS- 04-R ____

____ Princeton ___________________________________ ___NJ _ ___ Mercer ___________ __08541___ - _0000___
City State County ZIP Code + 4

2. Applicant's D-U-N-S Number (b)(2) [redacted]

3. Applicant's T-I-N |_X_|_X_|-|_X_|_X_|_X_|_X_|_X_|_X_|_X_|

4. Catalog of Federal Domestic Assistance #: **84.**__3_|_0__|_5__|_K__|

Title: _ U.S. Dept. of Education – Institute of _______

_ Education Sciences _______________________

5. Project Director:___ Dylan Wiliam _________________________

Address:__Educational Testing Service, Rosedale Rd., MS -10-R

_Princeton___________________ __NJ__ __08541__ _______
City State Zip code + 4

Tel. #: (609) _734___-_1307___ Fax #: (609)_734___-_1755___

E-Mail Address: __ dwiliam@ets.org ___________________

6. Novice Applicant ___Yes _X_No

7. Is the applicant delinquent on any Federal debt? ___Yes _X_No
(If "Yes," attach an explanation.)

8. Type of Applicant *(Enter appropriate letter in the box.)* |_I__|

A - State	F - Independent School District
B - Local	G - Public College or University
C - Special District	H - Private, Non-profit College or University
D - Indian Tribe	I - Non-profit Organization
E - Individual	J - Private, Profit-Making Organization

K - Other (*Specify*): ______________________________________

Application Information

9. Type of Submission:

-PreApplication
___ Construction
___ Non-Construction

-Application
___ Construction
_X_ Non-Construction

10. Is application subject to review by Executive Order 12372 process?

___ Yes (*Date made available to the Executive Order 12372 process for review)*: ____/____/_________

_X_ No *(If "No," check appropriate box below.)*
_X__ Program is not covered by E.O. 12372.
___ Program has not been selected by State for review.

11. Proposed Project Dates: _9___/_1__/_2004___ _6__/_30_/_2008____
Start Date: **End Date:**

12. Are any research activities involving human subjects planned at any time during the proposed project period?
_X_ Yes (Go to 12a.) ___ No (Go to item 13.)

12a. Are **all** the research activities proposed designated to be exempt from the regulations?
__ Yes (Provide Exemption(s) #): ____

_X_ No (Provide Assurance #): ___XXX XXXXXXXX________

13. Descriptive Title of Applicant's Project:

____ Developing and Using Diagnostic Items in ________

____ Mathematics and Science ___________________

Estimated Funding

14a. Federal	**$ _1392034________. 00**	
b. Applicant	**$ ________________. 00**	
c. State	**$ ________________. 00**	
d. Local	**$ ________________. 00**	
e. Other	**$ ________________. 00**	
f. Program Income	**$ ________________. 00**	
g. TOTAL	**$ __1392034_______. 00**	

Authorized Representative Information

15. To the best of my knowledge and belief, all data in this preapplication/application are true and correct. The document has been duly authorized by the governing body of the applicant and the applicant will comply with the attached assurances if the assistance is awarded.

a. Authorized Representative (*Please type or print name clearly.*)

__Patricia R. Beckman____

b. Title: **__ Director, Corporate Contracts & Intellectual Property Management**

c. Tel. #: (609) __734______-_1683_______ Fax #: (609) _734____-__5183______

d. E-Mail Address: ___ pbeckman@ets.org____

e. Signature of Authorized Representative

____

Instructions for Form ED 424

1. **Legal Name and Address.** Enter the legal name of applicant and the name of the primary organizational unit which will undertake the assistance activity.

2. **D-U-N-S Number.** Enter the applicant's D-U-N-S Number. If your organization does not have a D-U-N-S Number, you can obtain the number by calling 1-800-333-0505 or by completing a D-U-N-S Number Request Form. The form can be obtained via the Internet at the following URL: **http://www.dnb.com.**

3. **Tax Identification Number.** Enter the taxpayer's identification number as assigned by the Internal Revenue Service.

4. **Catalog of Federal Domestic Assistance (CFDA) Number.** Enter the CFDA number and title of the program under which assistance is requested. The CFDA number can be found in the federal register notice and the application package.

5. **Project Director.** Name, address, telephone and fax numbers, and e-mail address of the person to be contacted on matters involving this application.

6. **Novice Applicant.** Check **"Yes"** or "**No**" only if assistance is being requested under a program that gives special consideration to novice applicants. Otherwise, **leave blank.**

   Check "**Yes"** if you meet the requirements for novice applicants specified in the regulations in 34 CFR 75.225 and included on the attached page entitled "Definitions for Form ED 424." By checking "Yes" the applicant certifies that it meets these novice applicant requirements. Check "**No**" if you do not meet the requirements for novice applicants**.**

7. **Federal Debt Delinquency.** Check **"Yes"** if the applicant's organization is delinquent on any Federal debt. (This question refers to the applicant's organization and not to the person who signs as the authorized representative. Categories of debt include delinquent audit disallowances, loans and taxes.) Otherwise, check **"No."**

8. **Type of Applicant.** Enter the appropriate letter in the box provided.

9. **Type of Submission.** See "Definitions for Form ED 424" attached.

10. **Executive Order 12372.** See "Definitions for Form ED 424" attached. Check **"Yes"** if the application is subject to review by E.O. 12372. Also, please enter the month, day, and four (4) digit year (e.g., 12/12/2001). Otherwise, check **"No."**

11. **Proposed Project Dates.** Please enter the month, day, and four (4) digit year (e.g., 12/12/2001).

12. **Human Subjects Research.** (See I.A. "Definitions" in attached page entitled "Definitions for Form ED 424.")

    If Not Human Subjects Research. Check "**No**" if research activities involving human subjects are not planned at any time during the proposed project period. The remaining parts of Item 12 are then not applicable.

    If Human Subjects Research. Check "**Yes**" if research activities involving human subjects are planned at any time during the proposed project period, either at the applicant organization or at any other performance site or collaborating institution. Check "**Yes**" even if the research is exempt from the regulations for the protection of human subjects. (See I.B. "Exemptions" in attached page entitled "Definitions for Form ED 424.")

12a. **If Human Subjects Research is Exempt from the Human Subjects Regulations.** Check "**Yes**" if all the research activities proposed are designated to be exempt from the regulations. Insert the exemption number(s) corresponding to one or more of the six exemption categories listed in I.B. "Exemptions." In addition, follow the instructions in II.A. "Exempt Research Narrative" in the attached page entitled "Definitions for Form ED 424." Insert this narrative immediately following the ED 424 face page.

12a. **If Human Subjects Research is Not Exempt from Human Subjects Regulations.** Check "**No**" if some or all of the planned research activities are covered (not exempt). In addition, follow the instructions in II.B. "Nonexempt Research Narrative" in the page entitled "Definitions for Form ED 424." Insert this narrative immediately following the ED 424 face page.

12a. **Human Subjects Assurance Number.** If the applicant has an approved Federal Wide (FWA) or Multiple Project Assurance (MPA) with the Office for Human Research Protections (OHRP), U.S. Department of Health and Human Services, that covers the specific activity, insert the number in the space provided. If the applicant does not have an approved assurance on file with OHRP, enter "None." In this case, the applicant, by signature on the face page, is declaring that it will comply with 34 CFR 97 and proceed to obtain the human subjects assurance upon request by the designated ED official. If the application is recommended/selected for funding, the designated ED official will request that the applicant obtain the assurance within 30 days after the specific formal request.

Note about Institutional Review Board Approval. ED does not require certification of Institutional Review Board approval with the application. However, if an application that involves non-exempt human subjects research is recommended/selected for funding, the designated ED official will request that the applicant obtain and send the certification to ED within 30 days after the formal request.

13. **Project Title.** Enter a brief descriptive title of the project. If more than one program is involved, you should append an explanation on a separate sheet. If appropriate (e.g., construction or real property projects), attach a map showing project location. For preapplications, use a separate sheet to provide a summary description of this project.

14. **Estimated Funding.** Amount requested or to be contributed during the first funding/budget period by each contributor. Value of in-kind contributions should be included on appropriate lines as applicable. If the action will result in a dollar change to an existing award, indicate **only** the amount of the change. For decreases, enclose the amounts in parentheses. If both basic and supplemental amounts are included, show

breakdown on an attached sheet. For multiple program funding, use totals and show breakdown using same categories as item 14.

15. **Certification.** To be signed by the authorized representative of the applicant. A copy of the governing body's authorization for you to sign this application as official representative must be on file in the applicant's office. Be sure to enter the telephone and fax number and e-mail address of the authorized representative. Also, in item 15e, please enter the month, day, and four (4) digit year (e.g., 12/12/2001) in the date signed field.

Paperwork Burden Statement. According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless such collection displays a valid OMB control number. The valid OMB control number for this information collection is **1875-0106**. The time required to complete this information collection is estimated to average between 15 and 45 minutes per response, including the time to review instructions, search existing data resources, gather the data needed, and complete and review the information collection. **If you have any comments concerning the accuracy of the estimate(s) or suggestions for improving this form, please write to:** U.S. Department of Education, Washington, D.C. 20202-4651. **If you have comments or concerns regarding the status of your individual submission of this form write directly to:** Joyce I. Mays, Application Control Center, U.S. Department of Education, 7th and D Streets, S.W. ROB-3, Room 3633, Washington, D.C. 20202-4725

Definitions for Form ED 424

Novice Applicant (See 34 CFR 75.225). For discretionary grant programs under which the Secretary gives special consideration to novice applications, a novice applicant means any applicant for a grant from ED that—

- Has never received a grant or subgrant under the program from which it seeks funding;
- Has never been a member of a group application, submitted in accordance with 34 CFR 75.127-75.129, that received a grant under the program from which it seeks funding; and
- Has not had an active discretionary grant from the Federal government in the five years before the deadline date for applications under the program. For the purposes of this requirement, a grant is active until the end of the grant's project or funding period, including any extensions of those periods that extend the grantee's authority to obligate funds.

In the case of a group application submitted in accordance with 34 CFR 75.127-75.129, a group includes only parties that meet the requirements listed above.

Type of Submission. "Construction" includes construction of new buildings and acquisition, expansion, remodeling, and alteration of existing buildings, and initial equipment of any such buildings, or any combination of such activities (including architects' fees and the cost of acquisition of land). "Construction" also includes remodeling to meet standards, remodeling designed to conserve energy, renovation or remodeling to accommodate new technologies, and the purchase of existing historic buildings for conversion to public libraries. For the purposes of this paragraph, the term "equipment" includes machinery, utilities, and built-in equipment and any necessary enclosures or structures to house them; and such term includes all other items necessary for the functioning of a particular facility as a facility for the provision of library services.

Executive Order 12372. The purpose of Executive Order 12372 is to foster an intergovernmental partnership and strengthen federalism by relying on State and local processes for the coordination and review of proposed Federal financial assistance and direct Federal development. The application notice, as published in the Federal Register, informs the applicant as to whether the program is subject to the requirements of E.O. 12372. In addition, the application package contains information on the State Single Point of Contact. An applicant is still eligible to apply for a grant or grants even if its respective State, Territory, Commonwealth, etc. does not have a State Single Point of Contact. For additional information on E.O. 12372 go to http://www.cfda.gov/public/eo12372.htm.

PROTECTION OF HUMAN SUBJECTS IN RESEARCH

I. Definitions and Exemptions

A. Definitions.

A research activity involves human subjects if the activity is research, as defined in the Department's regulations, and the research activity will involve use of human subjects, as defined in the regulations.

—Research

The ED Regulations for the Protection of Human Subjects, Title 34, Code of Federal Regulations, Part 97, define research as "a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge." *If an activity follows a deliberate plan whose purpose is to develop or contribute to generalizable knowledge it is research.* Activities which meet this definition constitute research whether or not they are conducted or supported under a program which is considered research for other purposes. For example, some demonstration and service programs may include research activities.

—Human Subject

The regulations define human subject as "a living individual about whom an investigator (whether professional or student) conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information." *(1) If an activity involves obtaining information about a living person by manipulating that person or that person's environment, as might occur when a new instructional technique is tested, or by communicating or interacting with the individual, as occurs with surveys and interviews, the definition of human subject is met. (2) If an activity involves obtaining private information about a living person in such a way that the information can be linked to that individual (the identity of the subject is or may be readily determined by the investigator or associated with the information), the definition of human subject is met.* [Private information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a school health record).]

B. Exemptions.

Research activities in which the **only** involvement of human subjects will be in one or more of the following six categories of ***exemptions*** are not covered by the regulations:

(1) Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (a) research on regular and special education instructional strategies, or (b) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (a) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (b) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation. ***If the subjects are chil-***

dren, exemption 2 applies only to research involving educational tests and observations of public behavior when the investigator(s) do not participate in the activities being observed. Exemption 2 does not apply if children are surveyed or interviewed or if the research involves observation of public behavior and the investigator(s) participate in the activities being observed. [Children are defined as persons who have not attained the legal age for consent to treatments or procedures involved in the research, under the applicable law or jurisdiction in which the research will be conducted.]

(3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior that is not exempt under section (2) above, if the human subjects are elected or appointed public officials or candidates for public office; or federal statute(s) require(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.

(4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

(5) Research and demonstration projects which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine: (a) public benefit or service programs; (b) procedures for obtaining benefits or services under those programs; (c) possible changes in or alternatives to those programs or procedures; or (d) possible changes in methods or levels of payment for benefits or services under those programs.

(6) Taste and food quality evaluation and consumer acceptance studies, (a) if wholesome foods without additives are consumed or (b) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the Food and Drug Administration or approved by the Environmental Protection Agency or the Food Safety and Inspection Service of the U.S. Department of Agriculture.

II. Instructions for Exempt and Nonexempt Human Subjects Research Narratives

If the applicant marked "Yes" for Item 12 on the ED 424, the applicant must provide a human subjects "exempt research" or "nonexempt research" narrative and insert it immediately following the ED 424 face page.

A. Exempt Research Narrative.

If you marked "Yes" for item 12 a. and designated exemption numbers(s), provide the "exempt research" narrative. The narrative must contain sufficient information about the involvement of human subjects in the proposed research to allow a determination by ED that the designated exemption(s) are appropriate. The narrative must be succinct.

B. Nonexempt Research Narrative.

If you marked "No" for item 12 a. you must provide the "nonexempt research" narrative. The narrative must address the following seven points. Although no specific page limitation applies to this section of the application, be succinct.

(1) **Human Subjects Involvement and Characteristics**: Provide a detailed description of the proposed involvement of human subjects. Describe the characteristics of the subject population, including their anticipated number, age range, and health status. Identify the criteria for inclusion or exclusion of any subpopulation. Explain the rationale for the involvement of special classes of subjects, such as children, children with disabilities, adults with disabilities, persons with mental disabilities, pregnant women, prisoners, institutionalized individuals, or others who are likely to be vulnerable

(2) **Sources of Materials**: Identify the sources of research material obtained from individually identifiable living human subjects in the form of specimens, records, or data. Indicate whether the material or data will be obtained specifically for research purposes or whether use will be made of existing specimens, records, or data.

(3) **Recruitment and Informed Consent**: Describe plans for the recruitment of subjects and the consent procedures to be followed. Include the circumstances under which consent will be sought and obtained, who will seek it, the nature of the information to be provided to prospective subjects, and the method of documenting consent. State if the Institutional Review Board (IRB) has authorized a modification or waiver of the elements of consent or the requirement for documentation of consent.

(4) **Potential Risks**: Describe potential risks (physical, psychological, social, legal, or other) and assess their likelihood and seriousness. Where appropriate, describe alternative treatments and procedures that might be advantageous to the subjects.

(5) **Protection Against Risk**: Describe the procedures for protecting against or minimizing potential risks, including risks to confidentiality, and assess their likely effectiveness. Where appropriate, discuss provisions for ensuring necessary medical or professional intervention in the event of adverse effects to the subjects. Also, where appropriate, describe the provisions for monitoring the data collected to ensure the safety of the subjects.

(6) **Importance of the Knowledge to be Gained**: Discuss the importance of the knowledge gained or to be gained as a result of the proposed research. Discuss why the risks to subjects are reasonable in relation to the anticipated benefits to subjects and in relation to the importance of the knowledge that may reasonably be expected to result.

(7) **Collaborating Site(s)**: If research involving human subjects will take place at collaborating site(s) or other performance site(s), name the sites and briefly describe their involvement or role in the research.

Copies of the Department of Education's Regulations for the Protection of Human Subjects, 34 CFR Part 97 and other pertinent materials on the protection of human subjects in research are available from the Grants Policy and Oversight Staff, Office of the Chief Financial Officer, U.S. Department of Education, Washington, D.C. 20202-4248, telephone: (202) 708-8263, and on the U.S. Department of Education's Protection of Human Subjects in Research Web Site at http://www.ed.gov/offices/OCFO/humansub.html

Developing and using diagnostic items in mathematics and science

A proposal by Dr. Dylan Wiliam, Program Director, Educational Testing Service, for the Institute of Education Sciences

Contribution of the project to solving an educational problem

Competence in mathematics and science is essential for all. The world of work increasingly demands high levels of mathematical and scientific skill from employees, and recent studies have found that high-level mathematics and science skills confer greater lifetime salary benefits than high-level literacy skills (Wolf, 2002). However, the same skills also are increasingly important for the world outside work—for example, in terms of informed consumer choice and effective participation in democracy.

There is widespread agreement that the performance of U.S. students in grades K-12, while relatively better in science than mathematics, is modest by international comparisons (Gonzalez, Calsyn, Jocelyn, et al., 2000; Programme for International Student Assessment, 1999). While weaknesses in mathematics and science manifest themselves most markedly in high schools, these weaknesses have their origins much earlier. According to the Third International Mathematics and Science Study (TIMSS), the growth in mathematics achievement between fourth grade and eighth grade was smaller in the United States than in any of the other participating countries (Mullis, Martin, Beaton, Gonzalez, Kelly, & Smith, 1997) and while performance in science is better than it is in mathematics, the variation in science performance amongst 13-year-olds is larger in the United States than in any other country participating in TIMSS (Beaton, Martin, Mullis, Gonzalez, Smith, & Kelly, 1996, p. 26).

The reasons are, of course, complex and difficult to disentangle. Schmidt, McKnight, & Raizen (1997b) cite unfocused curricula and textbooks that give shallow coverage to a large number of topics —in their words, textbooks that are "a mile wide and an inch deep" (p. 62). Others have pointed to the lack of subject knowledge of U.S. teachers (Darling-Hammond, 1998; Ma, 1999), although there is little consensus about what kind of subject knowledge is necessary. There is some evidence that higher subject knowledge is important (Ferguson, 1991) although increasing subject knowledge produces diminishing returns in terms of student achievement (Anderson, 1991; Darling-Hammond, 2000). It therefore appears that thorough knowledge of the matter to be taught is more important than knowledge of the subject beyond what is to be taught (Askew, Brown, Rhodes, Johnson, & Wiliam, 1997; Ma, 1999).

Weaknesses in teacher subject knowledge may, over the long term, be helped by changes in teacher recruitment. When the economy weakens, the supply of well-qualified teachers tends to increase, but the converse is also true. When the economy strengthens, those who are well-trained in mathematics and science are precisely those who are most in demand (National Commission on Mathematics and Science Teaching for the 21st Century, 2000). For the foreseeable future, therefore, improving performance of young people in science and mathematics requires improving the performance of the teachers we already have, specifically through continuing professional development (Greenwald, Hedges, & Laine, 1996).

Important insights into the teaching of mathematics in particular are contained in the reports of the TIMSS 1999 video study, which showed that while, in the teaching of eighth grade mathematics, the United States differs radically from Japan, so do most other countries (Hiebert, Gallimore, Garnier, et al., 2003). In fact, on a range of indicators, practice in the United States lies within the range of what is found in other countries in terms of, for example, assignment of homework, use of goal and summary statements, procedural complexity of work given, relatedness of problems used, public presentation of solutions, and the use of calculators.

However, mathematics teaching in grade eight in the United States is distinctive in important ways. While U.S. teachers of mathematics dominate classroom talk less than the teachers in the other countries (an average of 8 teacher words for every student word, compared to 16 in Hong Kong and 13 in Japan), levels of student engagement are low. This points to the importance of looking at the quality, as well as the quantity of classroom interaction, and here, real differences between the United States and the other countries are apparent.

U.S. lessons are more fragmented than in the other five countries studied, with less development of themes, but U.S. lessons also are distinctive in that students spend much more time working individually on a large number of short exercises. When U.S. teachers do engage with the whole class the interaction is characterized by a series of low-order questions to which students give short answers, with the teacher assuming primary responsibility for the learning process—a result that Schmidt, Jorder, Cogan, et al. (1997a) found also applied to science classrooms. This is important, because, as Wang, Haertel, and Walberg (1993) found, the frequency and quality of teacher-student academic interactions were among the strongest correlates of positive student outcomes. In particular, they found that the kinds of questioning (high frequency, appropriate and challenging cognitive level, range of difficulty), the response types required (constructed response rather than selected response) and the provision of sufficient wait time (at least three seconds for higher-order questions) were strong predictors of academic success. Therefore just using a large number of questions is not enough. The questions need to be questions worth asking—questions that promote thinking and evoke evidence of understanding, which, when appropriately interpreted, provide guidance to teachers about what do to next.

In their review of research about *How people learn*, Bransford, Brown, and Cocking (2000) propose an eight point research agenda covering the development of educational materials:

1. Review a sample of current curricula, instructional techniques, and assessments for alignment with principles discussed in this volume
2. In the areas in which curriculum development has been weak, design and evaluate new curricula, with companion assessment tools, that teach and measure deep understanding
3. Conduct research on formative assessment
4. Develop and evaluate videotaped model lessons for broadly taught, common curriculum units that appear throughout the K-12 education system
5. Conduct extensive evaluation research through both small-scale studies and large-scale evaluations to determine the goals, assumptions, and uses of technologies in classrooms and the match or mismatch of these uses with the principles of learning and the transfer of learning
6. Conduct research on key conceptual frameworks, by discipline, for the units that are commonly taught in K-12 education

7. Identify and address preconceptions by field
8. Develop an interactive communications site that provides information on curricula by field.

This proposal directly addresses four of these goals (# 2, 3, 6, and 7) and indirectly addresses one more (# 4). Through the provision of diagnostic items, to be used in the course of day-to-day instruction and also in summative assessment, the project will provide assessment tools that that teach and measure deep understanding (goal 2). As well as providing models of high-quality questions that can be used by teachers to develop their own questioning, the questions will be directly applicable to the content standards of the majority of states.

By grounding the research in the emerging theory of formative assessment, and by developing our understanding of how teachers can use assessment tools in the service of learning, the project will expand our understanding both of what formative assessment is, and how teachers can be supported in developing this aspect of their instructional practice (goal 3). By providing conceptual frameworks for the chosen grades, the project will provide a resource that will be usable by teachers throughout the United States. At the same time, it will begin the process of integrating existing insights about student preconceptions with new insights obtained from analysis of large data-sets from constructed response items (goals 6 and 7). In addition, as part of the research effort, we will video-tape lessons in which teachers use the diagnostic items developed by the project, which should provide us with exemplar material for use in training other teachers (goal 4).

The project will represent three years of work. In the first year, we will review existing state standards for mathematics and science for one elementary school grade (grade 4) and one middle-school grade (grade 8) in order to identify the most commonly occurring topics. Then, we will review existing tests and analyze ETS data archives of student responses to constructed-response items to identify the most common errors made by students. Through careful reviewing of the research literature, these errors on individual items, and distinctive patterns of errors across items, will be related to misconceptions, and ideas for remediation will be developed. In the second year, we will work with small groups of teachers. We will introduce them to the items, share with them strategies for using these items in classrooms, and support them in doing so, while continuing to develop further items. In the third year, we will undertake a formal randomized trial with teachers, examining the impact on student achievement, focusing specifically on whether, as we expect, the incorporation of these items into teaching results in reduced within-class variance in achievement (Mevarech & Kramarski, 2003; White & Frederiksen, 1998).

The following section, which forms the bulk of this research narrative, outlines the background and context for the proposal, its relationship to existing research on formative assessment, the research questions to be addressed, and the details of how the research will be carried out.

Research Plan

Background and context

All 50 states have produced content standards for mathematics, although in one case (New York) this is combined with science and technology, and all but one state (Minnesota) have produced content standards for science, although again, these are sometimes combined with technology (Council of Chief State School Officers Division of State Services and Technical Assistance, 2003). Along with these state standards, many states also approve textbook series, which have been selected for alignment with the state's standards. Textbooks are then adopted at state, district, or school level, at significant cost, and, at least for mathematics, it is clear that most teachers build their lessons around commercially published textbooks and supplemental curricular materials (Ball & Cohen, 1996). The textbooks already in place therefore represent an "installed base" of curricular materials in which a great deal has been invested, both in terms of cash and in terms of the time that teachers have spent in adapting textbook resources for their own classrooms. It follows, therefore, that attempting to improve the achievement of students through changes in instructional materials will be costly, and slow. For this reason, for the short and medium term at least, finding ways of supplementing the existing "installed base" of curricular materials is likely to be the most effective way of improving student achievement in mathematics and science.

A large number of the texts currently in use promote a "standard operating procedure" of introducing a technique or idea, which is then followed by reinforcement through repetition. In this approach, all incorrect answers are treated as having the same information content, signaling that the necessary knowledge or skill has not been acquired. Although rarely stated, the theory of learning implied by such texts is "associationist" (Bransford et al., 2000), because failures to learn result in further reinforcement activities (in other words, it is implicitly assumed that the incorrect answer is caused by the failure in some associative link that needs to be strengthened).

The persistence of such underlying models of learning is all the more remarkable given that, over the last 50 years, there has been increasing consensus amongst psychologists of education that the learning of most mathematics and science is a constructive, rather than passive, activity (Driver, Squires, Rushworth, & Wood-Robinson, 1994; von Glasersfeld, 1991). From a constructivist perspective, all incorrect answers are not equivalent. Such errors as are made as the result of active processes of meaning-making on the part of the learner.

This was exploited in the 1970s and 1980s, in a series of research studies at the University of London, which explored ways in which assessments might support learning. The Concepts in Secondary Mathematics and Science (CSMS) project investigated mathematical and scientific reasoning in students through the use of tests that were intended to illuminate aspects of students' thinking, rather than just measure achievement (Hart, Brown, Kerslake, Küchemann, & Ruddock, 1985a; Shayer, 1981). Subsequent studies in mathematics developed detailed models of the strategies that students used (and the errors they made) when they responded to items in algebra, ratio and proportion, fractions and graphs (Booth, 1984; Hart, 1984; Kerslake, 1986; Sharma, 1993), and the way in which these strategies related to the instruction received (Johnson, 1989). In science, the work of Rosalind Driver (1983) and her collaborators has provided a rich source of students' conceptions in the elementary grades in biology (Osborne, Wadsworth, & Black, 1990b; Russell & Watt, 1990b), chemistry (Russell, Longden, & McGuigan, 1998a; Russell & Watt, 1990a), earth sciences (Osborne, Wadsworth, Black, & Meadows, 1990c; Russell, Bell, Longden, & McGuigan, 1993) and physics (Osborne, Black, Smith, & Meadows, 1990a; Osborne, Black, Smith, & Meadows, 1991; Russell, McGuigan, & Hughes, 1998b; Watt & Russell, 1990), and for the middle-school grades in biology (Bell &

Brook, 1984), chemistry (Brook, Briggs, & Driver, 1984b) and physics (Brook, Bell, Briggs, & Driver, 1984a; Brook & Driver, 1984).

While much subsequent research has expanded our knowledge of the conceptions that students bring to mathematics and science classrooms, these have tended to focus on the academic issues rather than how these insights can be used in practice, and no studies have used randomized trials to ascertain specifically whether the use of such items as a part of classroom practice does increase student achievement. The aim of the current project, then, is to draw together existing work on student conceptions in mathematics and science, to work with small groups of teachers in the United States to explore how diagnostic items can be integrated into existing teaching schemes, and to research the impact of the use of such items has on the overall levels, and variability, of student achievement.

Research basis for the proposal: Formative assessment

The term "formative evaluation" was first used, apparently, by Michael Scriven (1967) in connection with the curriculum and teaching, but it was Bloom, Hastings, and Madaus (1971) who were the first to use the term in its generally accepted current meaning. They defined *summative evaluation tests* as those tests given at the end of episodes of teaching (units, courses, etc) for the purpose of grading or certifying students, or for evaluating the effectiveness of a curriculum (p. 117). They contrasted these with "another type of evaluation which all who are involved—student, teacher, curriculum maker—would welcome because they find it so useful in helping them improve what they wish to do" (p. 117), which they termed "formative evaluation."

Reviews by Natriello (1987), Crooks (1988), and Black and Wiliam (1998) have provided clear evidence that improving the quality of formative assessment increases student achievement. Natriello's review covers the full range of assessment purposes (which he classified as certification, selection, direction, and motivation), while Crooks' review covered only what he termed "classroom evaluation." The review by Black and Wiliam built on four other key reviews of research published since those by Natriello and Crooks—reviews by Bangert-Drowns and the Kuliks into the effects of classroom testing (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991a; Bangert-Drowns, Kulik, & Kulik, 1991b; Kulik, Kulik, & Bangert-Drowns, 1990) and a review by Black of research on summative and formative assessment in science education (Black, 1993).

Natriello's review used a model of the assessment cycle, beginning with purposes, and moving on to the setting of tasks, criteria and standards, evaluating performance and providing feedback, and then discusses the impact of these evaluation processes on students. His most significant point was that the vast majority of the research in this area was largely irrelevant because of weak theorization, which resulted in key distinctions (e.g. the quality and quantity of feedback) being conflated. Crooks' paper had a narrower focus—the impact of evaluation practices on students. He concluded that the summative function of assessment has been too dominant and that more emphasis should be given to the potential of classroom assessments to assist learning. Most importantly, assessments must emphasize the skills, knowledge, and attitudes regarded as most important, not just those that are easy to assess. Black and Wiliam's review, like that of Crooks, focused specifically on day-to-day classroom assessment practices, and found that improvements in the quality of formative assessment resulted in effect sizes of the order of 0.4 to 0.7 standard deviations.

Taken together these reviews provide compelling evidence that improving the quality of formative assessment is a powerful mechanism for improving student achievement. The question is then whether these improvements have already been secured—after all, if all teachers are already doing this, then we can expect little improvement in student achievement by placing more emphasis on this aspect of practice. By its very nature, evidence about the extent of formative assessment practice is more difficult to find, and more limited than evidence about its effects, but such evidence as there is suggests that formative assessment is not well-developed in teaching. Bransford, Brown, and Cocking (2000) note that the knowledge-base on how to implement formative assessment effectively is weak (p. 257) and the evidence from the TIMSS video studies cited above suggests that few mathematics classrooms in the United States embody these ideas fully.

However, what the reviews cited above also reveal is that the term formative assessment is not well defined. For example, while the reviews by Natriello and Crooks cited 91 and 241 references respectively, only nine references were cited in both reviews. Bransford, Brown, and Cocking (2000) suggest that formative assessment involves "making students' thinking visible by providing frequent opportunities for assessment, feedback, and revision, as well as teaching students to engage in self-assessment" (p. 257). A review of other studies (Black & Wiliam, 1998; Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Duschl & Gitomer, 1997) suggest that the following are important aspects of formative assessment:

1. Higher-order, worthwhile questions and problems that promote thinking and learning
2. Deliberately designing in opportunities for students to express their understanding
3. Extending wait time after questions to allow students to think and formulate intelligent responses
4. Setting up the learning environment and classroom routines to engage all students, not just those who are "quick" and eager to show they know the "right answer"
5. Devising ways to efficiently analyze and understand the learning shown in student responses
6. Providing feedback to students that relates to the qualities of the work, identifies what the student can do to improve, and avoids comparisons with other students
7. Ensuring that students understand the goals of the work in which they are engaged
8. Training students in self-assessment so that they are able to monitor their progress towards their goals

From the foregoing, it can be seen that high-quality classroom questioning can help to raise achievement in two ways. The first is that if the questions are higher-order questions, requiring thought and extended responses from the students, then such questions can increase the degree of "mindfulness" (Bangert-Drowns et al., 1991a) with which the students engage in classroom discourse. Through questioning that promotes high-level discourse, connections are made that are vital in the development of understanding in both mathematics (Pugalee, 2001) and science (Otto & Schuck, 1983) and a range of other studies support the idea that high-level discourse is important for high achievement (Kilpatrick, Swafford, & Findell, 2001; Rand Mathematics Study Panel, 2003). The second way in which high-quality classroom questioning can help to raise achievement is by making the thinking of students visible, as noted by Bransford, Brown, and Cocking (*op cit*) so that the teacher and the students can take better-focused instructional measures to advance learning. These two aspects are subsumed by Perrenoud (1998) into a more general theoretical framework of the *regulation of learning processes*. He notes that in the interactive regulation of learning processes, there is a macro-level process of setting up "didactical situations" in which the conceptions of learners are evoked and the micro-level process of guiding learning in the light of the evidence generated by the didactical situations. Conceiving of formative assessment as the regulation of learning, rather than just feedback, incorporates all eight of the elements of formative assessment identified above, but also admits of other ways in which learning is kept "on track," for example when a teacher "does not intervene in person, but puts in place a 'metacognitive culture,' mutual forms of teaching and the organization of regulation of learning processes run by technologies or incorporated into classroom organization and management" (p. 100). The emphasis, at all times, is on ensuring that the processes in place guide students' learning towards fuller understanding of key mathematical and scientific concepts.

In order to regulate learning effectively, information is needed about the learner's current state of learning, the desired goal state, and the learning experiences that are likely to move the learner towards the goal. Put simply, there is a need to know where the learner is in their learning, where they are going, and what to do to get there. Crossing this three-fold typology of information needs with the different agents in the classroom (the student, her or his peers, and the teacher) creates the following framework for aspects of regulating learning.

	Where the learner is	Where they are going	How to get there
Teacher	Evoking information	Establishing goals	Feedback
Peer	Peer-assessment	Sharing success criteria	Peer-tutoring
Student	Self-assessment	Sharing success criteria	Self-directed learning

To establish where the learner is in their learning, the teacher needs a range of ways of evoking information, and it is this aspect of the regulation of learning that is the main focus of this proposal. However, other aspects also are important to the regulation of learning. For effective learning to take place, it also is necessary to be clear about the desired outcome of the learning. In some cases, this may be a particular goal (e.g., getting the students to be able to find the area of a trapezoid, or balance a chemical equation) but in the case of open-ended and exploratory work, there may be a whole range of goals that are appropriate for different learners or for learners at different stages of development. In the former case, the regulation of learning will be relatively tight, so that the teacher will attempt to "bring into line" all learners who are not heading towards the particular goal sought by the teacher. In other cases, particularly those involving open-ended and exploratory work, regulation will be relatively loose, with a teacher intervening to bring the learners "into line" only when the trajectory of the learner is radically different from that intended by the teacher. Once the outcomes are clear, the provision of feedback from the teacher can assist learning, provided, of course, such feedback is task-involving rather than ego-involving (Kluger & DeNisi, 1996). Learning also is enhanced when learners are able to assess their own performance (Fontana & Fernandes, 1994). But as Sadler (1989) notes, this requires that learners come to share the criteria for success that the teacher already has in mind—in other words, learners must become enculturated into the community of practice of which the teacher is already a full participant (Wiliam, 2001). Learners often find this difficult, however, and the involvement of peers can help learners understand success criteria and monitor their own progress towards their goals (White & Frederiksen, 1998). Thus peer-assessment provides an important complement to, and may even be a pre-requisite for, effective self-assessment (Black, Harrison, Lee, Marshall, & Wiliam, 2003).

Useful probes for evoking evidence of student achievement range in type from short-answer questions to elaborate problem situations. While many books on classroom questioning for teachers emphasize the importance of open-ended questions, closed-questions can be very effective when they relate to specific preconceptions that students hold. For example, asking students which of \F(3,7) and \F(3,11) is the larger is clearly a closed question, with only two answers, and only one of which is correct. Nevertheless, it is a useful item since many students believe that \F(3,11) is larger than \F(3,7) because 11 is larger than 7. In fact, in a survey of 14-year-olds in the United Kingdom, the facility of this item was only around 15%, despite the fact that around 75% of the students could correctly choose the larger of two ‘typical’ fractions like \f(2,3) and \F(3,4) (Hart, Brown, Kerslake, Küchemann, & Ruddock, 1985b). In elementary science, asking students whether grass is living or whether a truck is living are again both closed items, but are useful because they relate to well-known misconceptions about living things (i.e. that all—and only—living things move, while all non-living things are inanimate, and vice-versa).

At the other extreme, in both mathematics and science, well-chosen problem-situations can be highly illuminative of students’ thinking. This is well-established in Japanese education through the idea of a “big question” (*hatsumon*) at the beginning of an instructional sequence (Lewis, 2002). For example students can be asked what happens to the reading on a spring balance, which is suspending a weight inside a bell jar when the air is evacuated. While many students can answer this correctly, many others state that the reading on the spring balance will fall, because they believe that objects derive weight from the air pressure acting on them.

By themselves, however, the supply of questions is not enough. Attention also is needed to how the questions are used in the classroom. Surveys of practice in mathematics and science classrooms have consistently found that the amount of time that teachers allow students to begin their response to a question is very limited (Rowe, 1974; Tobin, 1987; Wragg & Brown, 2001). And yet it also is clear that students prefer classrooms in which they are given time to think (Altiere & Duell, 1991), and that they learn more in such classrooms (Tobin, 1984). However, it also is important to note that the effectiveness of questioning depends on the kinds of questions asked (Riley, 1986). Questioning students directly on their learning does improve retention of facts, but only when students are asked *about* their understanding is the retention of concepts and principles enhanced (Burton, Niles, Lalik, & Reed, 1986; Hamilton, 1985; King, 1990).

In all this, it also is important to note that the effects of questions cannot be understood independently of the context in which they are asked (Carlsen, 1991). The meaning of a question depends on the prevalent discourse of the classroom in which it is used, and the same initial prompt may function as a higher-order question in one classroom, but as a lower-order question in another, depending on the networks of expectations established between the teacher and the students—what Brousseau has called the ‘didactic contract’ (Brousseau, 1997).

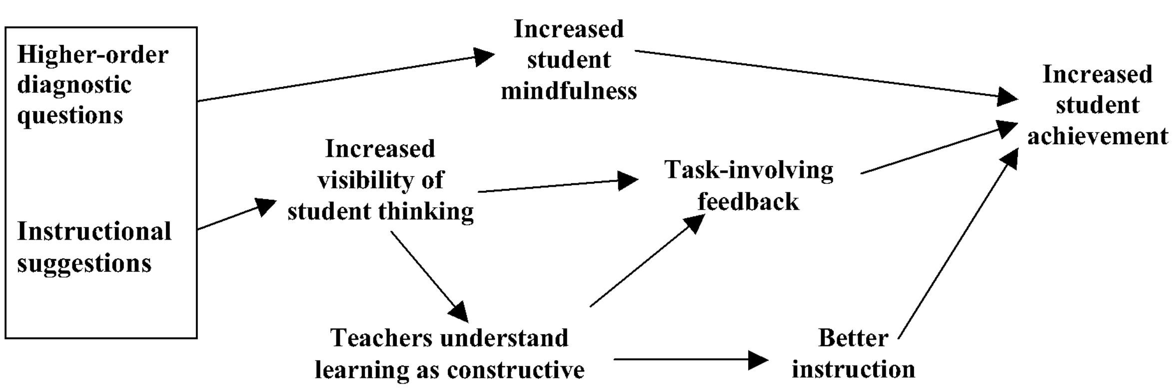
Research basis of the proposal: teacher professional development

There is increasing agreement that to be effective in raising student achievement, teacher professional development needs to attend to both *process* and *content* elements (Reeves, McCall, & MacGilchrist, 2001; Wilson & Berne, 1999). On the process side, professional development is more effective when it is related to the local circumstances in which the teachers operate (Cobb, McClain, Lamberg, & Dean, 2003), takes place over a period of time rather than being in the form of one-day workshops (Cohen & Hill, 1998), and involves teacher in active, collective participation (Garet, Birman, Porter, Desimone, & Herman, 1999). In addition to these process elements, however, professional development is more effective when it has a focus on deepening teachers' knowledge of the content they are to teach, the possible responses of students, and strategies that can be utilized to build on these (Supovitz, 2001).

Research questions

The major research question is whether providing teachers with diagnostic questions, suggestions for how to interpret and act on students' responses, and associated professional development, increases student achievement (Research Question 1).

The theory of the intervention is that providing teachers with high-quality questions will increase the "mindfulness" of students, which will in turn increase achievement (Bell, 1993). The same items will increase the "visibility" of students' thinking, which will support teachers in giving task-involving rather than ego-involving feedback which again will improve achievement (Kluger & DeNisi, 1996). Finally, it is expected that the increased visibility of student thinking will encourage teachers to understand the constructive nature of learning, leading to better instruction and thus to increased student achievement. The theory of the intervention is summarized in the following diagram:



Specifically, in the present study, we will address Research Question 1 by investigating four main research hypotheses:

> The mathematics achievement of fourth-grade students taught by teachers using diagnostic mathematics questions will be higher than that of fourth-grade students taught by comparable teachers not using such items.

The science achievement of fourth-grade students taught by teachers using diagnostic science questions will be higher than that of fourth-grade students taught by comparable teachers not using such items.

The mathematics achievement of eighth-grade students taught by teachers using diagnostic mathematics questions will be higher than that of eighth-grade students taught by comparable teachers not using such items.

The science achievement of eighth-grade students taught by teachers using diagnostic science questions will be higher than that of eighth-grade students taught by comparable teachers not using such items.

While these four questions are the *sine qua non* of the current proposal, we also are interested in understanding how teachers do, and do not, use diagnostic questions in classrooms, to help us make sense of the results, and also to help "sharpen" the intervention so that if, as we expect, the intervention is successful, it will be possible to spread the intervention more quickly without compromising its depth and sustainability (Coburn, 2003). This leads to the following subsidiary research questions:

- To what extent do teachers incorporate diagnostic questions into their teaching? (RQ2)
- How does the use of diagnostic items in teaching change classroom practice? (RQ3)

How the research will be carried out

The project will have three overlapping phases. In the first phase, starting September 2004 and going until April 2006, diagnostic items will be developed. In the second phase, from February 2005 to August 2006, we will work with small groups of elementary and middle school teachers, supporting them in developing their use of diagnostic items as part of their practice, and we will use this experience to revise the diagnostic questions and to develop training materials. In the third phase, from January 2006 to August 2007, a randomized trial will compare the performance of students taught by teachers trained in the use of the items with the performance of students taught by teachers who have not received either the items, or the associated professional development. In this final phase, we also will address the second and third research questions through classroom observations.

Phase 1: Item development

We think that it is important to work in both elementary and middle schools, in order to explore the differences between how subject specialists (in middle schools) and generalists (in elementary schools) use diagnostic items. In both elementary and middle schools, we propose to work with teachers in grades 4 and 8. There are several reasons for this choice. Grade 8 is typically the last year of middle school, and students by this point should have acquired most of the knowledge and skills needed for a postsecondary education. Grade 4 also represents a key point in the mathematics and science curriculum. In addition, these are the grades covered by the National Assessment of Educational Progress (see below), providing the project with a large number of publicly available open-ended items to which thousands of students have responded, thus removing the need for expensive data collection. In other words, we can use NAEP open-ended questions to identify modal student preconceptions and need not devote resources to collecting data on such error patterns.

During the initial months of the project, we will analyze the state content standards for mathematics and science for grades 4 and 8 and agree, with the Advisory Panel (see below) on the topics that are sufficiently common to warrant attention. Inevitably such a judgment will be somewhat subjective, and will need to be taken only when the analysis of the state standards has been completed, but at this stage, we envisage that topics will be included if they appear in the standards of 70% of the states for grades 4 or 8, or if they appear in the grade 4 or grade 8 standards of half the states and also appear in standards for 80% of the states for some grade. As a project deliverable, we will submit a report summarizing the review of state standards and identifying topics to be included among the diagnostic question set. This report will be submitted by December 2004.

Once the topics are identified, we will then begin the process of developing diagnostic items for each topic. ETS is ideally placed to do this work, since it has a huge range of both multiple-choice and constructed response items suitable for grade 4 and grade 8 mathematics and science. As discussed below, these items, while not originally developed to serve diagnostic purposes, nevertheless contain a wealth of information about student preconceptions, and can thus help guide the development process.

The number of items that we will develop for each topic will, of course, depend on the topic, but as a guide to the scope of the project, we envisage generating approximately 100 items for each of the two mathematics grades being studied in the project (i.e., grades 4 and 8) and 50 items for each of the two science grades. By combining these items with already available diagnostic items developed by others, such as the CSMS project (see Hart, 1985, and Appendix B) and the work of Driver and her colleagues cited above, we hope to create a bank of one good diagnostic question for each mathematics and science lesson in grade 4 and 8.

The actual development of the items will be undertaken by reflexive engagement with both theoretical insights and empirical data. In the first instance, for each topic, we will identify all relevant items at ETS that are available to the project. Then, we will review existing research literature on the preconceptions for that topic. For multiple-choice items, we will then look at the distractors for that item to determine whether the distractors are related to any of the preconceptions reported in the literature. If they are, then we will retrieve the response data for that item to determine the proportion of the response sample choosing that distractor to judge the prevalence of that preconception. While not all multiple-choice items will have distractors that are sufficiently well related to students' preconceptions, many will, and will therefore provide good starting points for the development of better items.

For constructed response items, we will retrieve the responses and response data, and begin classifying the responses to determine the most prevalent incorrect answers. For each item, we will retrieve at least 100 responses, and we will record any response that is given by at least 3% of the respondents. Our reason for choosing this threshold is that, to be useful to teachers, any preconception must be held by at least one student in each class, and we have assumed class sizes of 35. Because most of the multiple-choice items were developed for summative, rather than formative, purposes, we expect that most of the diagnostic items will be derived from the constructed-response items. Again by reference to the research literature, we will then "make sense" of the common preconceptions by developing models of student thinking that would lead to the preconception.

For example, one of the items developed by the CSMS project mentioned above asks: "If $e + f = 8$, then $e + f + g = ...$?" In the research sample of 12, 13, and 14-year-old students in England and Wales, 25%, 41% and 50% respectively gave the correct answer (i.e., "$8 + g$" or "$g + 8$"), but 29%, 26% and 16% respectively gave a numerical answer. More importantly, almost all of the students who gave numerical answers to this item gave one of three responses: 9, 12, or 15 (Hart, 1981). The students giving numerical responses appear to give the algebraic quantity a specific value in order to be able to work with it, but each of the three numerical values have a logic of their own. The students giving 9 as a response appear to be assuming that adding "g" to "$e + f$" must make it larger, and so choose the next integer after 8. Those giving 12 appear to have substituted a value of 4 for each of e, f, and g, and those giving 15 have presumably imputed values of 3 and 5 to e and f, and a value of 7 for g. This could be for a variety of reasons, but the reasoning most consistent with what is known about students' conceptions in algebra is that they believe that algebraic quantities are specific fixed unknowns, and that each letter must represent a different number (a common, and well-known consequence of certain instructional approaches). If e and f cannot be equal then the "obvious" choice is 3 and 5, and extending the arithmetic progression leads to a choice of 7 for g. Thus, in the process of item development, analysis of items will lead to new theories, and development of new theories will suggest new items, or families of items.

For topics that generate a number of different models, the next step will be to relate the models theoretically to each other. For example, Hart (1981) found six different categories of response to algebra items:

- letter evaluated
- letter not used
- letter used as an object
- letter used as a specific fixed unknown
- letter used as a generalized number
- letter used as a variable

Finally, for each item, and each incorrect response or category of responses, we will develop suggestions that can be made to teachers for instructional activities that would assist students in improving their understanding (Booth, 1984; Hart, 1981; Johnson, 1989).

All questions and materials developed will meet ETS standards for quality and fairness. They will all be reviewed for accuracy, editorial clarity, sensitivity and fairness, and grade appropriateness. Trained ETS assessment experts will conduct all the reviews.

The initial pool of diagnostic items will be developed initially at ETS and then reviewed by the members of our Advisory Panel. The initial set of items, for use in the pilot, will be reviewed in April 2005. Following this, the initial pool of items and related instructional materials will be produced into booklets, and submitted to the IES project officer for review by the end of May. Following review by the IES project officer, the booklets of items and related materials will be produced in forms appropriate to initial pilot testing. There will be further reviews in November 2005 and May 2006.

Phase 2: Piloting the implementation of items and related instructional materials with Teachers

Many studies, notably the Cognitively Guided Instruction (CGI) program (Carpenter, Fennema, Franke, Levi, & Empson, 2000), have found that students' learning is enhanced when teachers have detailed knowledge about the attainment of their students. However, for this knowledge to be effective, teachers also must have a range of instructional strategies to use this knowledge, because without such strategies, teachers will be wasting their time collecting information that they cannot use (Helmke & Schrader, 1987). For this reason, we do not believe that it will be beneficial just to send the items to teachers and see what they make of them. Rather the use of the items will be supported by a program of professional development that will be developed during the second year of the project, and used in the third year of the project as part of the randomized trial.

For the pilot phase of the project, we plan to work with 48 teachers—24 in middle schools and 24 in elementary schools. We will begin by identifying two clusters of middle schools and two clusters of elementary schools. Each middle school cluster will involve three middle schools (preferably in the same district) from each of which we will work with two mathematics and two science specialists, making a total of 12 eighth-grade teachers in a cluster—24 in all. Each elementary school cluster will involve six elementary schools, preferably from the same districts as the participating middle schools, and two teachers will be chosen from each participating elementary school, making a total of 12 fourth-grade teachers per cluster—again 24 in all.

We will not seek to make the pilot teachers representative of the profession as a whole, since during the pilot we are interested in understanding how teachers can best use these items. We therefore will recruit teachers who are "above average" in that they are already interested in developing their practice and have already begun to think seriously about their use of questions. Since an important part of the project's work involves teachers trying out diagnostic questions with their classes, we propose a pattern of meetings that involves six tightly focused half-day professional development workshops over the course of the pilot.

In order to minimize the costs of the venue for the six workshops with the teachers, and travel time for project staff, we will, wherever possible, hold workshops with middle-school teachers and elementary school teachers on the same day (e.g. meeting with middle-school teachers in the morning and elementary-school teachers in the afternoon). We have estimated the cost of substitute teachers as $200 per day (or $100 per teacher per meeting), which we have based on advice from school districts on what is required to ensure the availability of appropriate substitute teachers.

The meetings will follow the successful pattern of meetings for such interventions developed by Black, Harrison, Lee, Marshall, and Wiliam (2003, pp. 17-29). In the first two meetings, teachers will be introduced to the key concepts of formative assessment, and will be shown examples of good diagnostic items. Participants will then be shown examples of incorrect or incomplete student responses to the items and encouraged to develop models of student thinking that would lead to the production of such responses. Also in the first two meetings, the participating teachers will begin the process of "action planning" for changes to their practice (Harrison, 2000). In the third and subsequent meetings, teachers will meet and discuss their progress with their colleagues (approximately half the time will be spent in plenary sessions, and half the time in subject-specific groupings). They will be introduced to strategies of classroom organization that can be used to maximize the opportunity for peer-assessment and peer-tutoring (Black et al., 2003). In the intervention, particular emphasis will be given to the fact that many misconceptions are widely shared as noted above, and thus it is not necessary for teachers to individualize instruction for all students, but rather that they can use the responses to questions to determine temporary within-class groupings for instruction.

We expect the meetings to take place in April, June, September, and November 2006 and February and April 2007 (see project timeline in Appendix A). Within two weeks of each set of meetings, we will submit a brief report summarizing the key results of the meetings to the IES project officer. Following the completion of the pilot testing stage we will write a complete report summarizing the results of the field activities, and discussing the changes that will be made to the items and materials in the light of the pilot. This report will be submitted by August 2006. Finally, the materials will be revised and reproduced so that the next stage of the research—a randomized field trial—can proceed.

The training program is consistent with the emerging research on effective professional development outlined above. The research will be based in geographical clusters, allowing teachers to work collaboratively, and over an extended period of time, with others in their locality. Most importantly, the professional development focuses on student learning (and in particular, helping teachers to develop their own models of how students learn) and instructional decision-making, thus leading to more effective instruction.

Phase 3: Randomized trial

While there are undoubtedly problems in implementing randomized trials even in areas such as medicine (Collins & Pinch, 1998), there is growing agreement that such trials permit the strongest rejections of rival interpretations of outcome data, allowing the effect of the intervention to be isolated (Shavelson & Towne, 2002). In educational settings, such trials are particularly useful because the extent to which new approaches are taken up and used by teachers is often confounded with teacher quality (Cohen, Raudenbush, & Ball, 2002). For this reason, in its final year, the proposed project incorporates a randomized trial of the impact of the use of diagnostic questions by teachers on student achievement.

We will identify a total of 96 teachers to be involved in the experimental trial. Half of these (i.e., 48 teachers) will be eighth-grade teachers and half will be fourth-grade teachers. Three clusters each with four middle schools will be selected by negotiation with the district and the school. For each middle school, two mathematics and two science specialists will be selected. Ideally this would be a random assignment, but since the main randomization in this design is at school level (see below), where random selection of teachers is not acceptable to the school, we will attempt, for each school, to produce a representative grouping of teachers, for example by selecting one experienced and one recently qualified teacher (Black et al., 2003). Of course, the statistical power of the design (see below) could theoretically be increased by choosing only one teacher from each of 96 schools, but this would isolate the teachers in the school, reducing the support that they get from colleagues. This would weaken the effectiveness of the intervention, thus reducing the effect size and the power. Each elementary school cluster, as in Phase 2, will involve six elementary schools, preferably from the same districts as the participating middle schools, and two teachers will be chosen from each participating elementary school. The design assumes that the participating elementary school teachers are not subject specialists, but to minimize the burden, participating teachers will be given items for either mathematics or science, and to provide mutual support, both teachers in the same school will be using the same items (i.e., mathematics or science)

In the selection of districts, schools and teachers for the trial, we will first select a sample of districts within New Jersey, Delaware, and Pennsylvania, through negotiation with district superintendents. At this stage we envisage that each of the districts will be large enough to have four middle schools, so a total of three school districts will be needed to generate the 12 schools and 48 teachers needed for the middle-school sample. If it is not possible to find three districts with four middle schools, then clusters of districts will be formed to generate the 12 schools needed. At that stage in the selection process (January 2006), the principals of the schools will be approached and invited to collaborate with the project although they will not be told whether they will be in the experimental group or the control group. The incentive for the experimental group will be the professional development provided as part of the experimental trial during 2006-2007. The incentive for the control group will be that they will be given access to the items at the end of the trial in August 2007, and given professional development in their use (at ETS's expense) beginning September 2007. Two schools in each district (i.e., six in total) will then be selected for the intervention, with the other two in each district (i.e., six in total) acting as controls.

For elementary schools, a similar process will be followed, but due to the smaller size of many elementary schools, and the fact that each school will be participating for a single subject, we will select only two teachers from each school, so a total of 24 schools will be required—eight schools in each district or cluster. Within each district or cluster of eight schools four will be selected at random to be experimental schools, and, again at random, two will be designated as mathematics schools and two as science schools. The arrangement of the allocation of treatments to teachers is as shown in the table below.

	Experiment			Control		
	Math	Science	General	Math	Science	General
Grade 4	12	12				24
Grade 8	12	12		12	12	

In order to measure the increases in mathematics and science attainment of the students taught by teachers participating in the trial, four tests will be created from released multiple-choice NAEP items, one for fourth-grade mathematics, one for eighth-grade mathematics, one for fourth-grade science and one for eighth-grade science. Two forms of each test will be prepared giving a total of eight test forms (M4A, M4B, S4A, S4B, M8A, M8B, S8A, S8B). Each class will be given one form of the appropriate test at the beginning of the trial (in September 2006) and the other form at the end of the trial in June 2007). In order to keep the administration of the tests simple, all students in the same class will receive the same form at the same time, so for example, twelve of the eighth-grade mathematics classes (six experimental and six control classes) will take form A of the eighth-grade mathematics test in September 2006 and form B in June 2007, while the other twelve classes will take form B of the eighth-grade mathematics test in September 2006 and form A in June 2007.

The degree of randomization could be further increased by "spiraling" the test forms so that at each administration, half the students in a participating class would be given one form and half of the class the other. However, to accomplish this with a test-retest design, this would require teachers to ensure that each student received a different form in June 2007 from the one they received in September 2006. Our experience is that such designs are frequently too complicated to be administered reliably, with the result that some students receive the same form of the test twice, impairing the integrity of the design. Since teachers are treated as a random rather than a fixed effect in this study, little will be lost by not spiraling the test forms.

The intervention for the experimental teachers will be based on that used in Phase 2, modified in the light of our findings, involving six half-day workshops. Because part of the expected impact of the intervention is by increasing the subject knowledge, information on the teachers' subject knowledge at the beginning and the end of the trial would be useful. However, our experiences in negotiating with schools and districts to-date suggest that any requirement for teachers to be assessed on their subject knowledge would be a significant disincentive to participation, and in particular might skew the sample by being a greater disincentive to those who perceive their own subject knowledge as limited. Since the main aim of the trial is to produce robust data on the impact of the intervention on the achievement of students, we will not administer tests of subject knowledge to the participating teachers. However, we will ask all participants to complete questionnaires about their attitudes and beliefs about mathematics or science and teaching in June 2006 and towards the end of the project in May 2007.

During Phase 3 we also will visit the classrooms of the participating teachers to see how their practice is changing over time. We have considered using available classroom observation protocols, but these are insufficiently focused on formative assessment for the needs of this project. We will therefore adapt the observation protocol developed in the United Kingdom by the King's-Medway-Oxfordshire Formative Assessment Project (Lee, 2001) for use in U.S. classrooms. Where practically feasible, and consent from students and teachers is forthcoming, we will video lessons taught by the participating teachers for use in developing further training materials. Teachers participating in the workshops will be asked to keep a journal of their experiences. These journal logs will provide rich sources of information about the reflections of the teachers in the project, but will also provide very valuable for use with subsequent cohorts of teachers, in "scaling up" the project, if it is successful.

Data analysis

While we are primarily interested in the impact of the intervention on the achievement of students, the unit of analysis here must be the teacher. By comparing test outcomes between treated and untreated student in a randomized trial, we can be reasonably sure that if there are significant differences in progress, they can be attributed to the use of the diagnostic questions and the training in their use. It could, of course, be argued that this is not a fair trial since the experimental teachers also will have been given professional development opportunities that the control teachers did not get. This would be a valid criticism if one of our research questions was "Does giving teachers diagnostic items improve student achievement," but we do not believe that these items can be used effectively by teachers without some professional support.

The primary form of analysis will be ANOVA with gain scores as the dependent variable and treatment as the independent variable (and this is the basis on which we have estimated the power of the design—see below). However, in order to investigate the robustness of the results, and to investigate whether, as hypothesized, the intervention reduces the variability of achievement, we also will use ANCOVA, with post-test scores as the dependent variable, pre-test scores as a covariate, and treatment as the independent variable.

The qualitative data, from classroom observations, will be used to address research questions 2 and 3. Data analysis will begin with reviewing of videotaped lessons and lesson observations forms, bearing in mind the theoretical underpinnings of the research (i.e., the regulation of learning). This will lead to the generation of categories from the data (Glaser & Strauss, 1967; Strauss, 1987). We will search for discrepant cases in the data and sample purposively for these as the study progresses as a way of further refining the developing theory.

Ethical and security considerations

ETS accepts responsibility to safeguard research information from unauthorized access, disclosure, modification, or destruction. The Committee for Prior Review of Research is responsible for ensuring that the rights of human participants in ETS research projects are properly protected. Using an established process, all research projects are reviewed for compliance with the guidelines related to human subjects stated in the *ETS Standards for Quality and Fairness.* In addition, the American Psychological Association's *Ethical Principles in the Conduct of Research with Human Participants* and the Department of Health and Human Services' regulations on protection of human subjects (45 CFR 46 – Code of Federal Regulations) also provide substantive guidance for the review process.

This review process ensures that the research is of minimal risk to the participant and that proper procedures are being implemented to protect the individual right to privacy. As regards privacy, all response protocols and any information about individuals and institutions are identified by code only. Information linking code numbers to names is kept in a separate secure location. The linkage is kept only as long as necessary for coding and follow-up studies, and then the record of actual names is destroyed. The release of data to secondary sources, or a use of data not originally intended, requires special review. In those cases where the data are individually identifiable, additional participant consent also is required. ETS adheres to guidelines compliant with the Privacy Act of 1974, and is committed to the confidentiality of its work concerning human subjects.

The process also entails providing a candid, understandable description of the potential subjects' research tasks, and of the treatment and use of the data. This description is the basis for the participating individual, or parents of individuals under 18, to consent to participate. Obtaining consent is a requirement of the research.

The organizations with which ETS works are encouraged to adopt policies and procedures that adequately protect the confidentiality of the data transferred to those organizations to and from ETS.

Statistical Power of the Design

To ensure that we have a good chance of being able to detect a statistically significant difference, we have conducted power analyses for the proposed design. We have assumed that the students are taught in classes of average size 20 students, that the scores on the first administration and the second administration are correlated at 0.75, that the intra-class correlation is 0.2, and that the standard deviation in gain scores from first to second administration will be 70% of the standard deviation of scores the second administration.

On this basis the probability that an effect-size of 0.3 will generate a statistically significant result (with a one-tailed test, and $\alpha = 0.05$) is 0.87, and, of course, the chance to detect will be higher if the effect is larger (Bloom, 2003; Cohen, 1988). Our estimate of the minimum effect size is based on the evidence from other studies on the effectiveness of formative assessment (Mevarech & Kramarski, 2003; Wiliam, Lee, Harrison, & Black, to appear, 2004).

Project deliverables

At the end of the project, there will be several tangible products in addition to the reports described above. The first will be evidence about the effects of the use of high-quality diagnostic items on student achievement. The second tangible product will be a set of approximately 600 diagnostic items (400 mathematics and 200 science) that are closely tied to the content standards of the states, and therefore ready to use in almost any elementary or middle school classroom. The third tangible product will be a training package that can be used by states and districts to begin the process undertaken in this project in their own region.

Personnel

Dr. Dylan Wiliam, the Director of the Learning and Teaching Research Center at ETS, will be the director of this project. Dr. Wiliam is internationally acknowledged as an authority on formative assessment, and has published more than 150 books, journal articles, and chapters in edited works on education. He has taught both mathematics and science in public schools in the United Kingdom, has been involved extensively in the pre-service and in-service education of schoolteachers and was, for five years, Dean of the School of Education at King's College London (the fifth largest university in the United Kingdom).

In order to exploit ETS's resources to the fullest, Dr. Wiliam will work with two senior ETS staff who will act as directors for the mathematics and science work.

Dr. Jody Underwood, a Development Scientist at ETS, will lead the mathematics work. Dr. Underwood has a background in computer science, cognitive science, and education and has particular expertise in mathematics education. Before coming to ETS she managed the NSF-funded ESCOT project, building applets for middle school mathematics and developed and piloted an online professional development course for teachers and for pre-service teachers focusing on teaching methods based on constructivist views of learning, with a special emphasis on formative assessment.

Jerry DeLuca, who is currently Director of Development at ETS for the National Board of Professional Teaching Standards (NBPTS), will lead the science work. Prior to this appointment, he was responsible for the development of science assessments for the program. He holds undergraduate and graduate degrees in science and has more than 20 years of successful classroom experiences as a science teacher, receiving several awards for teaching excellence. He has been closely involved with the NBPTS, being elected to its Board in 1996 and serving as the liaison for the development of the AYA (high school) assessment.

As well as their subject expertise, the three key staff members between them bring complementary skills in the area of formative assessment, cognitive science, and teacher professional development that will provide firm theoretical foundations for the project's work. To undertake the majority of the day-to-day work of the project, ETS will appoint a senior research associate, and the team will meet weekly during the first phase of the project, and all will take lead roles in the teacher workshops during both the pilot and trial phases. In addition, ETS will appoint a part-time staff assistant to support the work of the team.

Resources

Items

As the world's leading assessment organization, ETS has access to an unrivalled range of test items, and data on students' performance on those items. In some cases, the items are the property of ETS. Other items are owned by other bodies, but ETS has permission to use the data provided individual test-takers are not identified, and for some items, specific permission must be sought from the owner for any secondary analysis to be undertaken. For the present project, we envisage that only items in the first and second category will be necessary, and therefore all permissions regarding the use of the necessary items have been obtained. For example, ETS is responsible for the administration of the National Assessment of Educational Progress (NAEP) and therefore has access to the NAEP items (both released and secure) and to the performance of national representative cohorts of students on these items. The currently available NAEP items are listed in Appendix A.

Schools

We have begun negotiations with several school districts with a view to being involved in this work. Because of the strong focus on student learning and teachers' professional development, there has been strong interest in this work, and a keen desire to be involved in some way. We are therefore confident that we will have no difficulty in finding schools to participate in this study.

Advisory Group

Upon the commencement of the project, we will establish an advisory group of experts in the field of diagnostic assessment in mathematics and science education. The purpose of this group is to provide an external critical perspective on the work of the project, and to ensure that all relevant research is incorporated into the diagnostic items, and the associated recommendations for action to be taken. The group will have six experts external to ETS (three with expertise in mathematics education and three with expertise in science education). This group will advise the project on the selection of state content standards to be selected for coverage, and will be consulted on the diagnostic items to be used in the project.

Because of the nature of the project, with the majority of item development taking place within the first year, the advisory group will hold four one-day meetings in the first year, and one one-day meeting in each of the following two years.

We judge that, although they would normally command a much higher daily consultancy fee, the experts that we will invite to join the advisory group will be willing to participate for the standard maximum NSF consultancy rate of $500 per day, due to the interest and relevance of the project. In addition, we will meet all their travel costs (coach class air fares), which we have budgeted at $1000 per trip (to include transfers to and from airports) and the cost of meals and any necessary overnight accommodation (which we have assumed to be $200 per day). We estimate the cost of an appropriate venue for each meeting, with audio-visual support to be $1000.

References

Altiere, M. A., & Duell, O. K. (1991). Can teachers predict their students' wait time preferences? *Journal of Research in Science Teaching,* **28**(5), 455-461.

Anderson, L. W. (1991). *Increasing teacher effectiveness*. Paris, France: UNESCO International Institute for Educational Planning.

Askew, M.; Brown, M. L.; Rhodes, V.; Johnson, D. C., & Wiliam, D. (1997). *Effective teachers of numeracy: final report*. London, UK: King's College London School of Education.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher,* **25**(9), 6-8.

Bangert-Drowns, R. L.; Kulik, C.-L. C.; Kulik, J. A., & Morgan, M. T. (1991a). The instructional effect of feedback in test-like events. *Review of Educational Research,* **61 (2)**, 213 - 238.

Bangert-Drowns, R. L.; Kulik, J. A., & Kulik, C.-L. C. (1991b). Effects of frequent classroom testing. *Journal of Educational Research,* **85**(2), 89-99.

Beaton, A. E.; Martin, M. O.; Mullis, I. V. S.; Gonzalez, E. J.; Smith, T. A., & Kelly, D. L. (1996). *Science achievement in the middle school years: IEA's third international mathematics and science study*. Chestnut Hill, MA: Boston College.

Bell, A. W. (1993). Some experiments in diagnostic teaching. *Educational Studies in Mathematics,* **24**(1), 115-137.

Bell, B., & Brook, A. (1984). *Aspects of secondary students' understanding of plant nutrition: full report*. Leeds, UK: University of Leeds Centre for Studies in Science and Mathematics Education.

Black, P.; Harrison, C.; Lee, C.; Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Buckingham, UK: Open University Press.

Black, P. J. (1993). Formative and summative assessment by teachers. *Studies in Science Education,* **21**(1), 49-97.

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice,* **5**(1), 7-73.

Bloom, B. S.; Hastings, J. T., & Madaus, G. F. (Eds.). (1971). *Handbook on the formative and summative evaluation of student learning*. New York, NY: McGraw-Hill.

Bloom, H. S. (2003). *Sample design for an evaluation of the Reading First program*. Washington, DC: United States Department of Education Institute of Education Sciences.

Booth, L. R. (1984). *Algebra: children's strategies and errors*. Windsor, UK: NFER-Nelson.

Bransford, J. D.; Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brook, A.; Bell, B.; Briggs, H., & Driver, R. (1984a). *Aspects of secondary students' understanding of heat: full report*. Leeds, UK: University of Leeds Centre for Studies in Science and Mathematics Education.

Brook, A.; Briggs, H., & Driver, R. (1984b). *Aspects of secondary students' understanding of the particulate nature of matter: full report*. Leeds, UK: University of Leeds Centre for Studies in Science and Mathematics Education.

Brook, A., & Driver, R. (1984). *Aspects of secondary students' understanding of energy: full report*. Leeds, UK: University of Leeds Centre for Studies in Science and Mathematics Education.

Brousseau, G. (1997). *Theory of didactical situations in mathematics* (N. Balacheff, M. Cooper, R.Sutherland, V. Warfield, Trans.). Dordrecht, Netherlands: Kluwer.

Burton, J. K.; Niles, J. A.; Lalik, R. M., & Reed, W. M. (1986). Cognitive capacity engagement during and following interspersed mathemagenic questions. *Journal of Educational Psychology,* **78**(2), 147-152.

Carlsen, W. S. (1991). Questioning in classrooms: a sociolinguistic perspective. *Review of Educational Research,* **61**(2), 157-178.

Carpenter, T. P.; Fennema, E.; Franke, M. L.; Levi, L., & Empson, S. B. (2000). *Children's mathematics: cognitively guided instruction*. Portsmouth, NH: Heinemann.

Carpenter, T. P.; Fennema, E.; Peterson, P. L.; Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: an experimental study. *American Educational Research Journal,* **26**(4), 499-531.

Cobb, P.; McClain, K.; Lamberg, T. d. S., & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher,* **32**(6), 13-24.

Coburn, C. (2003). Rethinking scale: moving beyond numbers to deep and lasting change. *Educational Researcher,* **32**(6), 3-12.

Cohen, D. K., & Hill, H. C. (1998). *State policy and classroom performance: mathematics reform in California*. Philadelphia, PA: University of Pennsylvania Consortium for Policy Research in Education.

Cohen, D. K.; Raudenbush, S. W., & Ball, D. L. (2002). Resources, instruction, and research. In F. W. Mosteller & R. Boruch (Eds.), *Evidence matters: randomized trials in education research* (pp. 80-119). Washington, DC: Brookings Institution.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, H. M., & Pinch, T. (1998). *The golem at large: what you should know about technology*. Cambridge, UK: Cambridge University Press.

Council of Chief State School Officers Division of State Services and Technical Assistance (2003). *Map of standards by state: Web links to state standards*. Washington, DC: Council of Chief State School Officers.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research,* **58**(4), 438-481.

Darling-Hammond, L. (1998). Teachers and teaching: testing policy hypotheses from a National Commission report. *Educational Researcher,* **27**(1), 5-15.

Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education,* **51**, 166-173.

Driver, R. (1983). *The pupil as scientist?* Milton Keynes, UK: Open University Press.

Driver, R. H.; Squires, A.; Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science: research into children's ideas*. London, UK: Routledge.

Duschl, R. A., & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment,* **4**(1), 37-73.

Ferguson, R. (1991). Paying for public education: new evidence on how and why money matters. *Harvard Journal of Legislation,* **28**, 465-498.

Fontana, D., & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology,* **64**(4), 407-417.

Garet, M. S.; Birman, B. F.; Porter, A. C.; Desimone, L., & Herman, R. (1999). *Designing effective professional development: lessons from the Eisenhower Program*. Washington, DC: US Department of Education.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative research*. New York, NY: Aldine Publishing Company.

Gonzalez, P.; Calsyn, C.; Jocelyn, L.; Mak, K.; Kastberg, D.; Arafeh, S.; Williams, T., & Tsen, W. (2000). *Pursuing excellence: comparisons of international eighth grade mathematics and science achievement from a U.S. perspective, 1995 and 1999: initial findings from the Third International Mathematics and Science Study-Repeat*. Washington, DC: National Center for Education Statistics.

Greenwald, R.; Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research,* **66**(3), 361-396.

Hamilton, R. J. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research,* **55**(1), 47-85.

Harrison, C. (2000, September) *Action plans for formative assessment (work of the KMOFAP group).* Paper presented at the symposium, Getting Inside the Black Box : Formative Assessment in Practice, at the British Educational Research Association 26th annual conference held at Cardiff University. London, UK: King's College London School of Education.

Hart, K. M. (Ed.) (1981). *Children's understanding of mathematics: 11-16*. London, UK: John Murray.

Hart, K. M. (1984). *Ratio: children's strategies and errors*. Windsor, UK: NFER-Nelson.

Hart, K. M.; Brown, M. L.; Kerslake, D.; Küchemann, D., & Ruddock, G. (1985a). *Chelsea diagnostic mathematics tests: specimen set*. Windsor, UK: NFER-Nelson.

Hart, K. M.; Brown, M. L.; Kerslake, D.; Küchemann, D., & Ruddock, G. (1985b). *Chelsea diagnostic mathematics tests: teachers' guide*. Windsor, UK: NFER-Nelson.

Helmke, A., & Schrader, F. W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education,* **3**(2), 91-98.

Hiebert, J.; Gallimore, R.; Garnier, H.; Bogard Givvin, K.; Hollingsworth, H.; Jacobs, J.; Miu-Ying Chui, A.; Wearne, D.; Smith, M.; Kersting, N.; Manaster, A.; Tseng, E.; Etterbeek, W.; Manaster, C.; Gonzales, P., & Stigler, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study*. Washington, DC: National Center for Education Statistics.

Johnson, D. C. (Ed.) (1989). *Children's mathematical frameworks 8-13: a study of classroom teaching*. Windsor, UK: NFER-Nelson.

Kerslake, D. (1986). *Fractions: children's strategies and errors*. Windsor, UK: NFER-Nelson.

Kilpatrick, J.; Swafford, J. O., & Findell, B. (Eds.). (2001). *Adding it up: helping children learn mathematics*. Washington, DC: National Academy Press.

King, A. (1990). Enhancing Peer Interaction and Learning in the Classroom Through Reciprocal Questioning. *American Educational Research Journal,* **27**(4), 664-687.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention Theory. *Psychological Bulletin,* **119**(2), 254-284.

Kulik, C.-L. C.; Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: a meta-analysis. *Review of Educational Research,* **60**(2), 265-299.

Lee, C. (2001, September) *Trying to define what it is that makes a lesson formative.* Paper presented at British Educational Research Association 27th annual conference held at University of Leeds. London, UK: King's College London School of Education.

Lewis, C. C. (2002). *Lesson study: a handbook of teacher-led instructional change*. Philadelphia, PA: Research for Better Schools.

Ma, L. (1999). *Knowing and teaching elementary mathematics: teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mevarech, Z. R., & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology,* **73**(4), 449-471.

Mullis, I. V. S.; Martin, M. O.; Beaton, A. E.; Gonzalez, E. J.; Kelly, D. L., & Smith, T. A. (1997). *Mathematics achievement in the primary school years: IEA's third international mathematics and science study*. Chestnut Hill, MA: Boston College.

National Commission on Mathematics and Science Teaching for the 21st Century (2000, September). *Before it's too late: a report to the nation from the National Commission on Mathematics and Science Teaching for the 21st Century*. Report prepared for Secretary of Education. Washington, DC: United States Department of Education.

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist,* **22**(2), 155-175.

Osborne, J.; Black, P. J.; Smith, M., & Meadows, J. (1990a). *Light.* Liverpool, UK: University of Liverpool Press.

Osborne, J.; Black, P. J.; Smith, M., & Meadows, J. (1991). *Electricity*. Liverpool, UK: University of Liverpool Press.

Osborne, J.; Wadsworth, P., & Black, P. J. (1990b). *Processes of life*. Liverpool, UK: University of Liverpool Press.

Osborne, J.; Wadsworth, P.; Black, P. J., & Meadows, J. (1990c). *Earth in space*. Liverpool, UK: University of Liverpool Press.

Otto, P. B., & Schuck, R. F. (1983). The effect of a teacher questioning strategy training program on teaching behavior, student achievement, and retention. *Journal of Research in Science Teaching,* **20**(6), 521-528.

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning. Towards a wider conceptual field. *Assessment in Education: Principles Policy and Practice,* **5**(1), 85-102.

Programme for International Student Assessment (1999). *Measuring student knowledge and skills: a new framework for assessment*. Paris, France: Organisation for Economic Co-operation and Development.

Pugalee, D. K. (2001). Algebra for all: the role of technology and constructivism in an algebra course for at risk students. *Preventing School Failure,* **45**(4), 171-176.

Rand Mathematics Study Panel (2003). *Mathematical proficiency for all students: toward a strategic research and development program in mathematics education*. Santa Monica, CA: Rand Coroporation.

Reeves, J.; McCall, J., & MacGilchrist, B. (2001). Change leadership: planning, conceptualization and perception. In J. MacBeath & P. Mortimore (Eds.), *Improving school effectiveness* (pp. 122-137). Buckingham, UK: Open University Press.

Riley, J. P. (1986). The effects of teachers' wait-time and knowledge comprehension questioning on science achievement. *Journal of Research in Science Teaching,* **23**(4), 335-342.

Rowe, M. B. (1974). Wait time and rewards as instructional variables, their influence on language, learning and fate control. *Journal of Research in Science Teaching,* **11**, 81-94.

Russell, T.; Bell, D.; Longden, K., & McGuigan, L. (1993). *Rocks, soil and weather*. Liverpool, UK: University of Liverpool Press.

Russell, T.; Longden, K., & McGuigan, L. (1998a). *Materials*. Liverpool, UK: University of Liverpool Press.

Russell, T.; McGuigan, L., & Hughes, A. (1998b). *Forces*. Liverpool, UK: University of Liverpool Press.

Russell, T., & Watt, D. (1990a). *Evaporation and condensation*. Liverpool, UK: University of Liverpool Press.

Russell, T., & Watt, D. (1990b). *Growth*. Liverpool, UK: University of Liverpool Press.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science,* **18**, 119-144.

Schmidt, W. H.; Jorde, D.; Cogan, L. S.; Barrier, E.; Gonzalo, I.; Moser, U.; Shimizu, K.; Sawada, T.; Valverde, G. A.; McKnight, C. C.; Prawat, R. S.; Wiley, D. E.; Raizen, S. A.; Britton, E. D., & Wolfe, R. G. (1997a). *Characterizing pedagogical flow: an investigation of mathematics and science teaching in six countries*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Schmidt, W. H.; McKnight, C. C., & Raizen, S. A. (1997b). *A splintered vision: an investigation of U.S. science and mathematics education*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Scriven, M. (1967). *The methodology of evaluation.* Washington, DC: American Educational Research Association.

Sharma, S. S. (1993). *Graphs: children's strategies and errors*. London, UK: King's College London Centre for Educational Studies.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education.* Washington, DC: National Academy Press.

Shayer, M., & Adey, P. S. (1981). *Towards a science of science teaching: cognitive development and curriculum demand.* London, UK: Heinemann Educational Books.

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press.

Supovitz, J. A. (2001). Translating teaching practice into improved student achievement. In S. H. Fuhrman (Ed.) *From the capitol to the classroom: standards-based reform in the States* (pp. 81-98). Chcago, IL: University of Chicago Press.

Tobin, K. (1984). Effects of extended wait time on discourse characteristics and achievement in middle school grades. *Journal of Research in Science Teaching,* **21**(8), 779-791.

Tobin, K. (1987). The role of wait time in higher cognitive level learning. *Review of Educational Research,* **57**(1), 69-95.

von Glasersfeld, E. (Ed.) (1991). *Radical constructivism in mathematics education*. Dordrecht, Netherlands: Kluwer.

Wang, M. C.; Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research,* **63**(3), 249-294.

Watt, D., & Russell, T. (1990). *Sound.* Liverpool, UK: University of Liverpool Press.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition. Making science accessible to all students. *Cognition and Instruction,* **16**(1), 3-118.

Wiliam, D. (2001). An overview of the relationship between assessment and the curriculum. In D. Scott (Ed.) *Curriculum and assessment* (pp. 165-181). Greenwich, CT: JAI Press.

Wiliam, D.; Lee, C.; Harrison, C., & Black, P. J. (to appear, 2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles Policy and Practice,* **11**(2).

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: an examination of research on contemporary professional development. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (pp. 173-209). Washington, DC: American Educational Research Association.

Wolf, A. (2002). *Does education matter?* London, UK: Penguin.

Wragg, E. C., & Brown, G. (2001). *Questioning in the secondary school*. London, UK: RoutledgeFalmer.

Dylan Wiliam, Ph.D.

Dylan Wiliam is a senior researcher on the Evidence-Centered Teaching in Algebra and Its Impact on Student Learning. Dr. Wiliam received his Ph.D. from London University in 1993. He is the senior research director of the Learning & Teaching Research Center in ETS Research & Development in Princeton, NJ. He was a teacher of mathematics before joining King's College London in 1984, and he continued teaching at the college level until 2003, when he joined ETS. He served as dean and head of the School of Education at King's College for five years before advancing to assistant principal there. He is the author of more than 150 articles, books and papers on education, including several foundational pieces about formative assessment, some with Paul Black, his colleague at King's.

Dr. Wiliam has been able to combine a deep understanding of teachers, classrooms and schools with a rigorous approach to examining impact on student learning. At ETS, he provides R&D leadership in the area of formative assessment in schools, including requisite attention to the professional development of teachers.

Professional Preparation

Durham University	1976	B.S.	General Science (class III)
Open University (class I)	1983	B.A.	Mathematics & Science
South Bank Polytechnic		1985	M.S. Mathematics Education (Distinction) (CNAA)
London University	1993	Ph.D.	Education

Appointments

2003-present	Senior Research Director, Center for Learning & Teaching Research, Research & Development, Educational Testing Service, Princeton, NJ
2001-2003	Assistant Principal, King's College, University of London
1998-2003	Professor of Educational Assessment, King's College, University of London
1996-2003	Dean and Head of the School of Education, King's College, University of London
1994-1998	Senior Lecturer in Mathematics Education, King's College, University of London
1989-1991	Seconded as Academic coordinator to the Consortium for Assessment and Testing in Schools (CATS)
1986-1994	Lecturer in Mathematics Education, King's College, University of London
1984-1985	Research Fellow: Graded Assessment in Mathematics, Chelsea College, (later King's College, London).
1980-1984	Deputy head of Mathematics, scale 3, North Westminster school (ILEA); scale 4 from April 1982; acting head of department from September 1983.
1977-1980	Teacher of Mathematics, scale 1, Christopher Wren school (ILEA); scale 2 from January 1980.

1976-1977 Tutor in Mathematics and Physics, St. Cloud private residential VIth Form College.

Selected Publications

Wiliam, D., Black, P. J., Harrison, C., Lee, C., & Marshall, B. (2003). *Assessment for learning: putting it into practice*, pp. 160. Open University Press.

Wiliam, D. (2003). The impact of educational research on mathematics education. In A. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second International Handbook of Mathematics Education*, pp. 469-488. Dordrecht, Netherlands: Kluwer Academic Publishers.

Lester Jr, F. K., & Wiliam, D. (2002). On the purpose of mathematics education research: making productive contributions to policy and practice. In L. D. English (Ed.), *Handbook of International Research in Mathematics Education*, pp. 489-506. Mahwah, NJ: Lawrence Erlbaum Associates.

Wiliam, D., & Black, P. J. (2002). *Standards in public examinations,* pp. 48. King's College London Department of Education and professional studies.

Boaler, J., Wiliam, D., & Brown, M. L. (2000). Students' experiences of ability grouping – disaffection, polarization and the construction of failure. *British Educational Research Journal*, *27*(4), pp. 631-648.

Wiliam, D. (2000). The meanings and consequence of educational assessments. *Critical Quarterly*, *42*(1), pp. 105-127.

Wiliam, D. (1998). A framework for thinking about research in mathematics and science education. In J. A. Malone, W. Atweh, & J. R. Northfield (Eds.), *Research and Supervision in Mathematics and Science Education*, pp. 1-18. Mahwah, NJ: Lawrence Erlbaum Associates.

Collaborators

Black, P. J. – King's College London

Boaler, J. – Stanford University

Harrison, C. – King's College London

Lee, C. – King's College London

Lester Jr., F. K. – University of Indiana

Marshall, B. – King's College London

Jody S. (Gevins) Underwood

Jody Underwood is a Development Scientist in the Center for Assessment for Learning in the Product Development Division at ETS. Her research interests are in using technology in appropriate ways to facilitate learning, including evaluating learning technology to determine what helps and what detracts from learning, designing human-computer interfaces, as well as working with teachers to figure out how to deploy appropriate software in the classroom.

Dr. Underwood has a background in computer science, cognitive science, and education. She has focused on the learning of math, history, foreign languages, foreign alphabets, and literacy and has built online courses and tools in the service of learning. Before coming to ETS, she managed the NSF-funded ESCOT project to build applets for middle school mathematics and developed and piloted an online professional development course for preservice teachers focusing on formative assessment and constructivist teaching methods. In the last five years she has worked closely with middle-school teachers and students.

Professional Preparation

Hofstra University	1981	B.S.	Computer Science
Rutgers University	1987	M.S.	Computer Science (AI)
Technion-Israel Inst. of Technology	1999	Ph.D.	Science and Technology Education
Postdoctoral Institution: Vanderbilt University	1997-1999		Learning Technology Center

Professional Experience

2001 – present	Development Scientist, Center for Assessment for Learning, Product Development Division, Educational Testing Service, Princeton, NJ.
1999 – 2001	Educational Technologist, The Math Forum @ Drexel, (formerly of WebCT and of Swarthmore College), Swarthmore, PA
1997 – 1999	Postdoctoral Research Associate, Vanderbilt University, Postdoctoral position, Learning Technology Center, Nashville, TN.
1990 – 1993	Research Programmer, NASA Ames Research Center, Artificial Intelligence Research Branch, Moffett Field, CA.
1988 – 1990	Research Programmer, Carnegie Mellon University, Psychology Department, Pittsburgh, PA.
1984 – 1987	Research Programmer, Princeton University, Cognitive Science Laboratory, Princeton, NJ.
1984 – 1987	Teaching Assistant, Rutgers University, Computer Science Department, New Brunswick, NJ.
1981 – 1984	Programmer Analyst, Mt. Sinai Medical Center, Department of Biomathematical Sciences, New York City, NY.

Selected Publications

Hand, V., Roschelle, J., DiGiano, C., & Underwood, J. S. (2001). *Teaming with teachers to integrate technology and curriculum: Lessons from the ESCOT Project,* presented at the National Meeting of the American Educational Research Association, Seattle, WA.

Roschelle, J., & Underwood, J. S., Repenning, A., Jackiw, N., DiGiano, C., Alejandre, S. (2001). Producing interactive problems of the week: Component-based integration teams, presented at the National Educational Computing Conference, Chicago, IL.

Underwood, J.S., Weimar, S., Roschelle, J., & Barnes, D. (2000). *Working together on the use of technology in middle school mathematics: Perspectives of teachers, curriculum experts, and software developers in ESCOT,* presented at the National Meeting of the American Educational Research Association, New Orleans, LA.

Underwood, J.S. (1999) *Community building at the Math Forum,* presented at the Center for Innovative Learning Technology annual conference, San Jose, CA.

Underwood, J., Noser, T., Goldman, S.R. & Lawrence, S. (1998). *Cognitive aspects of the student inquiry process,* presented at the National Meeting of the American Educational Research Association.

Underwood, J. S. (1996). Data collection on the World Wide Web, *CÆLL Journal, 6* (4), pp. 13-20.

Baudin, C., Gevins, J. S., Baya, V., & Mabogunje, A. (1992). *Using domain concepts to index engineering design information,* Proceedings of the Conference of the Cognitive Science Society, Bloomington, Indiana.

Singley, M. K., Anderson, J. R., & Gevins, J. S. (1991). *Promoting abstract strategies in algebra word problem solving,* Proceedings of the International Conference on the Learning Sciences, Evanston, IL.

Singley, M. K., Anderson, J.R., Gevins, J. S., & Hoffman, D. (1989). *The algebra word problem tutor,* Proceedings of the 4th International Conference of AI and Education, Amsterdam, The Netherlands.

Synergistic Activities

- Discussant for the symposium: Multimedia Cases and Software Simulations for Teacher Learning, to be presented at the National Meeting of the American Educational Research Association, Seattle, WA, 2001.
- Presented a workshop: The Math Forum and ESCOT Present: Interactive Mentored Problems of the Week, at the National Conference for Teachers of Mathematics, Chicago, IL, 2000.
- Developed 'Web Librarian' – A book recommendation system to address reading problems in Middle School. A poster describing it was published in the Proceedings of the International Conference of the Learning Sciences, Atlanta, GA, 1998.

Collaborators & Other Affiliations

Collaborators: Janet Bowers, K. Ann Renninger, Wesley Shumar, Hollylynne Drier Stohl, Chris DiGiano, Jeremy Roschelle, Stephen Weimar

Graduate Advisor: Ehud Bar-On, InMotion, Ltd., Haifa, Israel

Postdoctoral Advisor: Susan Goldman, Vanderbilt University

Jerry L. DeLuca

With 23 years of classroom teaching experience in Secondary Science, Jerry DeLuca holds both undergraduate and graduate level degrees in Earth Science from California University of Pennsylvania. He has been recognized numerous times for his excellence in science teaching being named a Presidential Awardee for Excellence in Secondary Mathematics and Science Teaching, a GTE GIFT Fellow, an Ashland Teacher Achievement Award recipient, and a Milken Family Foundation National Educator. He served as the President of the West Virginia Science Teachers Association for two years and as Co-Principal Investigator for Project CATS (Coordinated and Thematic Science) a 5 year, five million dollar grant from the National Science Foundation to reform Science education statewide in West Virginia from 1994 – 1998.

In 1998, DeLuca joined ETS and currently is the Director of the National Board of Professional Teaching Standards (NBPTS) Program at the organization. DeLuca first became involved with the NBPTS when he was elected to its Board in 1996. He also was involved in the development of the NBPTS assessments for science teachers as he served as the liaison for the development of the AYA (high school) science assessment.

Professional Preparation

California University of Pennsylvania	1976	B.S.	Earth Science
California University of Pennsylvania	1981	M.S.	Earth Science

Fortyfive additional hours past Master's degree in science, mathematics, & education.
Credits were earned at the following institutions: California University of Pennsylvania, West Virginia University, Davis & Elkins College, Alderson Broaddus College, Salem College, and Glenville State College.

Professional Experience

2001-present	Director, National Board for Professional Teaching Standards Assessment Program
2000-2001	Director of Development, National Board for Professional Teaching Standards Assessment Program
1998-2000	Science Developer, National Board for Professional Teaching Standards Assessment Program
1981-1998	Science Teacher and Department Chair, Tucker County High School (TCHS), Tucker County Board of Education
1979-1981	Science Teacher, Parsons Elementary & Middle School
1978-1979	Science Teacher, Mountaineer High School
1977-1978	Substitute Teacher, Tucker County Schools

Selected Publications

DeLuca, J, various articles and President's Message, *Tucker County Education Association Newsletter.*

DeLuca, J. various articles and President's Message, *West Virginia Science Teachers Association Newsletter.*

DeLuca, J. Contributing author and editor, *West Virginia Science Curriculum Framework.*

DeLuca, J. "Social capitol", article for Appalachian Educational Laboratory - *The Link*, 1994

DeLuca, J. (Contributing editor) (1991) *Solid thinking about solid waste, an environmental curriculum for grades seven to nine*, Kraft General Foods Environmental Institute.

Leadership

Member, Board of Directors for the National Board for Professional Teaching Standards

Co-principal Investigator – Project CATS – A National Science Foundation Teacher Enhancement Grant, (five year, $4.1 million) to the state of West Virginia for statewide systemic science reform

Past President – West Virginia Science Teachers Association, 1993-1996

President Elect – West Virginia Science Teachers Association, 1991-1992

Cadre member – West Virginia Science Curriculum Framework

Distinguished Educator and Presenter – National Science Foundation Grant, Earth Science for West Virginia in the 21st Century

Chair – West Virginia Education Association State Legislative Committee, 1988 – 1996

West Virginia Education Association State Budget Committee, 1985 – 1988

Professional Memberships

National Science Teachers Association

West Virginia Science Teachers Association

National Education Association

BUDGET JUSTIFICATION

Below is a narrative explaining each of the budget line items on Ed. Form no. 524.

1. ED FORM 524 1. Personnel

Educational Testing Service Staff – Under the ETS accounting system, hourly employee salary rates are used to distribute the salary of staff to the various projects. [redacted] (b)(4)

The following charts detail the number of hours budgeted for ETS staff:

Year 1			PD	SRA	MD	SD	SA
2004	Sep	Analysis of state content standards	20	120	20	20	10
	Oct	Selection of core content	20	120	20	20	10
	Nov	Review of existing misconceptions literature	30	120	30	30	10
	Dec	Review of existing misconceptions literature	30	120	30	30	10
2005	Jan	Selection of core content items	10	40	10	10	20
	Jan	Error analysis	20	80	20	20	10
	Feb	Error analysis	30	110	30	30	10
	Feb	Identification of schools	30	110			25
	Mar	Error analysis	30	110	30	30	10
	Mar	Identification of schools	30	110			25
	Apr	Error analysis	30	110	30	30	10
	Apr	First meeting with pilot teachers	20	40	20	20	40
	May	Error analysis	30	120	30	30	10
	Jun	Second meeting with pilot teachers	20	40	20	20	40
	Jun	Error analysis	30	110	30	30	10
	Jul	Error analysis	30	110	30	30	10
	Aug	Error analysis	30	110	30	30	10
		Total	440	1680	380	380	270

PD: Project Director: Dylan Wiliam
SRA: Senior Research Associate (to be appointed)
MD: Mathematics Director: Jody Underwood
SD: Science Director: Jerry DeLuca
SA: Staff Assistant: Maria Rossi

Year 2			PD	SRA	MD	SD	SA
2005	Sep	Error analysis	20	85	20	20	10
	Sep	Third meeting with pilot teachers	20	40	20	20	40
	Oct	Error analysis	20	85	20	20	10
	Nov	Fourth meeting with pilot teachers	20	40	20	20	40
	Dec	Error analysis	30	125	30	30	10
2006	Jan	Error analysis	20	85	20	20	10
	Jan	Identification of schools to be involved in RCT	20	40	20	20	40
	Feb	Fifth meeting with pilot teachers	20	40	20	20	40
	Feb	Identification of schools to be involved in RCT	20	40	20	20	40
	Mar	Error analysis	30	125	30	30	10
	Apr	Sixth meeting with pilot teachers	20	40	20	20	40
	Apr	First meeting with RCT teachers	20	40	20	20	40
	Apr	Error analiysis	20	70	20	20	10
	May	Second meeting with RCT teachers	20	40	20	20	40
	Jun	Analysis of pilot data	20	70	20	20	10
	Jun	Collection of preliminary data	10	80	20	20	20
	Jul	Preparation of training materials for RCT	30	150	20	20	20
	Aug	Analysis of pilot data	30	150	20	20	20
		Total	390	1345	380	380	450

PD: Project Director: Dylan Wiliam
SRA: Senior Research Associate (to be appointed)
MD: Mathematics Dir.: Jody Underwood
SD: Science Director: Jerry DeLuca
SA: Staff Assistant: Maria Rossi

Year 3			PD	SRA	MD	SD	SA
2006	Sep	Compilation of pilot data	20	100	10	10	20
	Sep	Third meeting with RCT teachers	20	40	20	20	40
	Oct	Observation of experimental teachers & analysis	30	150	20	20	20
	Nov	Observation of experimental teachers & analysis	20	80	20	20	20
	Nov	Fourth meeting with RCT teachers	20	40	20	20	40
	Dec	Observation of experimental teachers & analysis	30	150	20	20	20
2007	Jan	Observation of experimental teachers & analysis	30	150	20	20	20
	Feb	Fifth meeting with RCT teachers	20	40	20	20	40
	Mar	Observation of experimental teachers & analysis	30	150	20	20	20
	Apr	Sixth meeting with RCT teachers	20	40	20	20	40
	May	Preparation of video materials	30	150	20	20	40
	Jun	Data collection and preliminary analysis	30	150	20	20	20
	Jul	Analysis and report writing	30	150	20	20	20
	Aug	Analysis and report writing	30	150	20	20	20
		Total	360	1540	270	270	380

PD: Project Director: Dylan Wiliam
SRA: Senior Research Associate (to be appointed)
MD: Mathematics Dir.: Jody Underwood
SD: Science Director: Jerry DeLuca
SA: Staff Assistant: Maria Rossi

Staff budget summary

	Year 1 Hours	Year 2 Hours	Year 3 Hours	Total Hours
PD	440	390	360	1,190
SRA	1,680	1,345	1,540	4,565
MD	380	380	270	1,030
SD	380	380	270	1,030
SA	270	450	380	1,100
Subtotal	3,150	2,945	2,820	8,915

The total amount budgeted for the staff hours listed above is as follows:

Year 1 $213,126
Year 2 $201,866
Year 3 $194,468
Total $609,460

2. ED FORM 524 2. Fringe Benefits

There are two components for cash fringe benefits. The first component is calculated by taking ten percent of the departmental charges. The second component is calculated by taking 32.5% of the base salary dollars.

Cash fringe benefits consist of:

- Legally required benefits which include FICA, Workmen's Compensation, Unemployment and Temporary Disability; and
- Insurance and retirement benefits, which include retirement annuity, group life, Major Medical, total disability, hospital-surgical and travel insurance for staff traveling on ETS business.

Fringe benefits are budgeted in the following way:

Year 1 $67,362
Year 2 $63,794
Year 3 $61,477
Total $192,633

3. ED FORM 524 3. Travel

This category consists of the following:

ETS Staff Travel - We have assumed that four members of ETS staff will travel by car to each pilot and trial meeting and will be reimbursed at the standard ETS rate of 35c per mile. We have assumed that these meetings will be no more than 60 miles from ETS, and that, for at least half the meetings it will be possible to hold two meetings on the same day.

Teacher Travel - We have assumed that the pilot and trial meetings will be close to the work-places of the participating teachers so that each teacher's travel cost to the meeting (at the ETS standard rate of 35c per mile) will be no more than $20. Since 48 teachers will each participate in six workshops, the total teacher travel budget for the pilot will be $5,760 ($1,920 falling in year 1 and $3,840 falling in year 2).

In the randomized trial (Phase 3), there will be three clusters of middle schools and three clusters of elementary schools so that each workshop will be delivered three times—a total of 36 workshops in all. Again we have assumed that the meetings will be close to the workplaces of the participating teachers so that each teacher's travel cost to the meeting will be no more than $20. Since 48 teachers will be participating in these meetings in the trial, the total cost will again be $5,760 ($1,920 falling in year 2 and $3,840 falling in year 3).

Advisory Panel Travel – There are six meetings planned (four in year 1 and one in each of years 2 and 3). Travel costs (coach class air fares) have been budgeted at $1,000 per trip (to include transfers to and from airports) and $200 has been allowed to cover the cost of meals and any necessary overnight accommodation.

4. ED FORM 524 4. Equipment
There are no charges in this category.

5. ED FORM 524 5. Supplies
This category consists of the following:

Telephone, copying, and office supplies budgeted at the amount of $1,000 per year for a total of $3,000.

6. ED FORM 524 6. Contractual
There are no charges in this category.

7. ED FORM 524 7. Construction
There are no charges in this category.

8. ED FORM 524 8. Other
This category consists of several components:

Honoraria - we have allowed an honorarium of $500 per day for the six members of the advisory panel for each of the six meetings planned (four in year 1 and one in each of years 2 and 3).

Production of Test Booklets - as mentioned in the research narrative, we will develop two parallel forms of each of four tests (mathematics and science at grades 4 and 8). Assuming that the elementary school students receive both the mathematics and science tests, we will need to produce and score approximately 6,000 test forms, which we have estimated will cost $8 each.

School Incentives - in providing the costs of substitute teachers for participating schools, we have estimated the cost as $200 per day (or $100 per teacher per half-day meeting). This is based on advice from local school districts on what is required to ensure the availability of appropriate substitute teachers. Since 48 teachers are involved in the pilot and 48 teachers are involved in the trial, the total cost of teacher–release is $57,600.

Meeting Expense - we have allowed $1,000 per meeting for the cost of an appropriate venue, including audio-visual support for the advisory panel meetings. For the pilot and trial meetings we have assumed that the cost of a venue for each half-day workshop will be $530.

Departmental Charges – Fifty-four percent of the departmental charges are included in this category.

10. ED FORM 524 10. Indirect Costs

General and Administrative Costs – General and administrative costs are computed at 16.26% of the budget items indicated and consist of those costs which cannot be assigned to a specific project but benefit all projects performed by ETS. Examples of these costs are those incurred by the Officers Division, Corporate Contracts and Proprietary Rights Office, Finance Division, the Library and the Human Resources Division. The ETS accounting system is audited by the Southern New Jersey Branch Office of the Mid-Atlantic Region of the Defense Contract Audit Agency.

Budgeted items that incur general and administrative costs are: ETS personnel, departmental charges, honoraria, booklet production, school incentives, advisory panel travel and teacher travel.

The dollar amounts for Indirect Costs are as follows:

Year 1 - $64,090
Year 2 - $59,899
Year 3 - $63,830
Total - $187,819

11. ED FORM 524 11. Training Stipends

There are no charges in this category.

Appendix A

Figures, Charts, or Tables That Supplement Research Text

Table A-1. **Developing and Using Diagnostics Items in Mathematics and Science: Indicative Project Timeline**

Table A-2. **TIMSS & PISA items available for use with the project**

Appendix B

Examples of Curriculum Material

Example B-1. **CSMS Algebra Items**

Example B-2. **CSMS Fractions Items**

Example B-3. **CSMS Ratio and Proportion Items**

Table A-1. Developing and Using Diagnostics Items in Mathematics and Science: Indicative Project Timeline

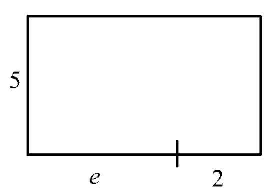
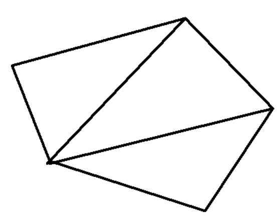
	2004				2005												2006												2007							
Activity	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A
Analysis of state standards	X																																			
Selection of core content		X																																		
Review preconceptions research			X	X																																
Error analysis					X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X																
Identification of pilot schools						X	X																													
Meetings with pilot teachers								X		X			X		X			X		X																
Identification of RCT schools																	X	X	X																	
Meetings with RCT teachers																				X		X			X		X			X		X				
Preparation of training materials																							X													
Observation of RCT teachers																										X	X	X	X	X	X					
Preparation of video materials																																X				
Data-collection																						X											X	X		
Analysis and report-writing																						X	X	X										X	X	X

Table A-2: TIMSS & PISA items available for use with the project

Objective	Grade	Status	MC	SCR	ECR
TIMSS 1995 Mathematics items					
Whole Numbers	3, 4	released	10	5	1
Fractions and Proportionality			6	2	4
Measurement and Number Sense			7	3	1
Data and Probability			4	2	2
Geometry			8	2	0
Patterns, Relations, and Functions			7	1	0
Fractions and Number Sense	7, 8	released	27	9	1
Algebra			13	3	2
Measurement			7	3	2
Geometry			16	1	0
Data and Probability			10	1	1
Proportionality			3	2	1
TIMSS 1999 Mathematics items					
Fractions and Number Sense	8	released	27	8	
Algebra			14	6	
Measurement			7	1	
Geometry			9		
Data and Probability			8	2	
TIMSS 1995 Science items					
Earth Science	3, 4	released	2	1	2
Life Science			20	5	3
Physical Science			15	4	3
Environment/Nature of Science			3	1	2
Earth Science	7, 8	released	6	5	
Life Science			18	9	
Chemistry			10	4	
Environment/Nature of Science			7	3	
TIMSS 1999 Science items					
Earth Science	8	released	10		
Life Science			25		
Chemistry			10		
Environmental/Resource Issues			6		
Scientific Inquiry/Nature of Science			4		
PISA 2000 Mathematics items					
Literacy	9, 10	released	5	4	2
PISA 2000 Science items					
Literacy	9, 10	released	4	3	1

Key: MC = multiple-choice, SCR = short constructed response, ECR = extended constructed-response

Example B-1. CSMS Algebra Items

no	question	wrong answers
1	Add 4 onto *n* + 5	*n* + 5 9
2	Add 4 onto 3*n*	7*n* 7
3	Multiply *n* + 5 by 4	4*n* + 5 *n* + 20 20
4	If e + f = 8 then e + f + g = . . .	15 12 8*g* 9
5	What is the area of this shape? 	*e* + 10 10 10*e* *e* + 7 7
6	*a* + 3*a* can be written more simply as 4*a*. Write this more simply, if possible: 2*a* +5*b*	7*ab* 12*a*
7	 In a shape like this, you can work out the number of diagonals by **taking away 3** from the number of sides. A shape with *k* sides has _______ diagonals.	8 9 *h* *i*

8	Mary's basic wage is £20 per week. She is also paid another £2 for each hour of overtime that she works. If h stands for the number of hours overtime that she works, and if W stands for her **total** wage (in £s), write down an equation connecting W and h.	$20 + 2h$ $W + 2h$ $20W + 2h$ $W = 10h$ $22 = W + h$ $W = 20 + h$ $20 + h$ $W + h$ $28 = W + 4h$
9	Blue pencils cost 5 pence each and red pencils cost 6 pence each. I buy some blue and some red pencils and altogether it costs me 90 pence. If b is the number of blue pencils bought, and if r is the number of red pencils bought, what can you say about b and r?	$b + r = 90$ $b + r$ $6b + 10r = 90$ $12b + 5r = 90$

Example B-2. CSMS Fractions Items

no	question	wrong answers
1	5 eggs in a box of 12 are found to be cracked.	
	(a) What fraction of the box of eggs is cracked?	$\frac{5}{7}, \frac{1}{5}, \frac{12}{5}$
	(b) What fraction of the box of eggs is not cracked?	$\frac{5}{7}, \frac{1}{7}, \frac{12}{7}$
2	Put these fractions in order of size, starting with the smallest. $\frac{1}{4}, \frac{1}{2}, \frac{1}{100}, \frac{1}{3}$	$\frac{1}{4}, \frac{1}{2}, \frac{1}{100}, \frac{1}{3}$ $\frac{1}{100}, \frac{1}{3}, \frac{1}{4}, \frac{1}{2}$ $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{100}$
3	Fill in the missing numbers:	
	(a) $\frac{1}{3} = \frac{2}{\ }$	$\frac{2}{4}, \frac{2}{3}$
	(b) $\frac{6}{8} = \frac{3}{\ }$	$\frac{3}{16}, \frac{3}{5}, \frac{3}{8}$
	(c) $\frac{5}{10} = \frac{\ }{30}$	$\frac{25}{30}, \frac{5}{30}, \frac{3}{30}$
	(d) $\frac{2}{3} = \frac{\ }{15}$	$\frac{14}{15}, \frac{2}{15}, \frac{5}{15}$
	(e) $\frac{4}{12} = \frac{1}{\ }$	$\frac{1}{9}, \frac{1}{12}, \frac{1}{4}$
4	$\frac{2}{7} = \frac{\square}{14} = \frac{10}{\bigcirc}$	
	a) What number goes in ❑?	1, 9, 6
	b) What number goes in ❍?	15, 28, 21
5	John pays $\frac{3}{5}$ of his wages in tax and $\frac{1}{10}$ of his wages on rent. What fraction of his wages does he have left after tax and rent have been paid?	$0, \frac{11}{15}, \frac{7}{10}, \frac{1}{5}$
6	A man is driving in France. He knows that 1 km is the same length as $\frac{5}{8}$ of a mile. His hotel is $\frac{2}{3}$ km from the petrol station. What is the distance in miles?	$\frac{7}{11}, \frac{1}{24}, \frac{31}{24}, \frac{16}{15}, \frac{15}{16}, \frac{12}{5}$

Example B-3. CSMS Ratio and Proportion Items

no	question	wrong answers
1	There are three eels, A, B and C in the tank at the Zoo. A 15cm long B 10cm long C 5cm long The eels are fed sprats, the number depending on their length.	
	(a) If C is fed 2 sprats, how many sprats should B and A be fed to match?	
	(i) B should be fed ____ sprats	3, 7, 12
	(ii) A should be fed ____ sprats	4, 8, 12, 17
	(b) If B eats 12 sprats, how many sprats should A be fed to match? A should be fed ____ sprats	13, 14, 17, 24, 27
	(c) If A gets 9 sprats, how many sprats should B get to match? B should be fed ____ sprats	4, $4\frac{1}{2}$, 7, 8, 19
2	Three other eels, X, Y and Z are fed with fishfingers, the length of the fishfingers depending on the length of the eel. Z 25cm long Y 15cm long X 10cm long	
	(a) If X has a fishfinger 2 cm long, how long should the fishfinger given to Z be?	4, 6, 8, 17, 27
	(b) If Y has a fishfinger 9 cm long, how long should the fishfinger given to Z be?	10, 11, 18, 19, 34

(c) If Z has a fishfinger 10 cm long, how long should the fishfingers given to X and Y be?

(i) X should get a fishfinger ____ cm long $2\frac{1}{2}$, 6, 8, 20

(ii) Y should get a fishfinger ____ cm long   5, 8, 9, 25