DOCUMENT RESUME

ED 393 158 EA 027 403

AUTHOR Falk, Beverly; And Others
TITLE An Invitation to Invention: Top-Down Support for Bottom-Up Reform of Assessment in New York State.
PUB DATE [96]
NOTE 30p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Competency Based Education; *Cooperative Learning; Educational Testing; *Evaluation Methods; High Schools; Holistic Evaluation; *Performance; *Student Evaluation; Testing Problems
IDENTIFIERS *New York

ABSTRACT

This paper describes the development and outcomes of New York State's "Invitation to Invention," the Regents Exam Options Project. Initiated as a temporary strategy to encourage and support local changes in teaching and assessment toward the learner-centered reforms taking place in the state, the project provided high school teachers and schools with temporary relief from some of the constraints of the current, highly structured state assessment system. The program invited them to create performance-based assessments (such as projects, performances, and portfolios) that would count toward a percentage of their students' high school Regents examination credit required for high school graduation. Specifically, the paper describes the impact of the first year of the project on teaching and learning and its potential for implementing change. Nine high schools and one school district participated. Data were derived from district reports, extensive sampling of student work, and interviews with and surveys of project teachers, students, and administrators. Findings indicate that the project changed the ways in which teachers taught, supported student achievement at higher performance levels, and enhanced teachers' professional learning. Concerns about conceptions of reliability and validity, the role of standardization, authenticity, and context in large-scale performance-based assessment systems are discussed. It is recommended that the state create performance assessments based on uniform standards in the disciplines--to use common scoring rubrics that articulate different levels of student performance. (Contains 27 references.) (LMI)

TM

An Invitation to Invention: Top-Down Support for Bottom-Up Reform of Assessment in New York State

Beverly Falk, National Center for Restructuring Education, Schools, and Teaching (NCREST)

and

Joanne Larson, New York State Education Department

with the assistance of Jianzhong Xu, NCREST

Mailing address:

Beverly Falk and Joanne Larson
NCREST, Box 110, Room 411
Room 411 Main Hall
Teachers College, Columbia University
New York, New York 10027
Phone: 212-678-3155
FAX: 212-678-4170

An Invitation to Invention: Top-Down Support for Bottom-Up Reform of Assessment in New York State

Abstract

This study describes the work of New York State's "Invitation to Invention" - The Regents Exam Options Project. Initiated as a temporary strategy to encourage and support local changes in teaching and assessment toward the learner-centered reforms taking place in the State, the project provided high school teachers and schools with temporary relief from some of the powerful and constraining aspects of the existing, highly-structured State assessment system. It invited them to create performance-based assessments (such as projects, performances, and portfolios embedded within the context of daily classroom learning) that would count toward a percentage of their students' high school Regents examination credit (required for high school graduation).

This study looks at the impact of the first year of this project - the effect it has had on teaching and learning as well as its consequences as a strategy for implementing change. It is based on detailed reports from the districts involved, extensive sampling of student work, interviews and surveys of all the project's teachers, students, and administrators.

Focusing on the consequences of authentic assessment development and use, this study describes how assessment can be a catalyst for changes in pedagogy, student and professional learning, as well as changes in the structures and culture of schools. It also raises questions about conceptions of reliability and validity, the role of standardization, authenticity, and context in large-scale performance-based assessment systems.

An Invitation to Invention: Top-Down Support for Bottom-Up Reform of Assessment in New York State

The last several decades have witnessed growing discrepancies between learning theory, teaching practices, and policies for testing. As our field has come to understand more about how people learn, how curriculum "coverage" can be the enemy of understanding, and how individuals possess multiple intelligences that need to be nurtured and responded to differently (Gardner,1983; Kornhaber and Gardner, 1993), many teachers have moved toward developing curricula that explore ideas in deeper ways than traditional approaches have heretofore allowed. They have developed learning experiences that engage students in studies and projects, that apply knowledge and information to real-life situations, and that provide opportunities for students to demonstrate higher-order skills and complex problem-solving abilities. Teaching strategies have also evolved that support students to learn in diverse ways and contexts.

As teachers and schools have moved in this direction they have found that traditional forms of assessment are not able to adequately reflect the more demanding and rigorous analytic skills nor the complex understandings their students are acquiring. This mismatch between curriculum and assessment has led to the development of a new genre of assessments that allow students to demonstrate what they know and can do in a variety of real-life tasks and situations. Inquiry and library research projects, portfolio collections of student work, journals, videos and oral tape recordings, science experiments and laboratory records are examples of these types of assessment (Darling-Hammond, et al., 1993). The information they yield is useful to teaching, enabling teachers to be diagnostic about students' learning so that they can develop curricula responsive to students' needs.

But teachers are limited in how far they can progress in the development of this work. Because in most school systems standardized tests still are the major determinants for high stakes decisions about students - promotion from grade, graduation, scholarships, etc. - they are powerful inhibitors to reform efforts aimed at

teaching and curriculum innovation. The press to "teach to the test" - to follow highly structured curricula that "cover" large bodies of content presented in much the same format of the test - makes it difficult to teach in ways that will ensure students' understandings and that will provide students with opportunities for inquiry, problem-solving, and in-depth exploration of their subjects. Testing pressure, in addition, makes it difficult to effectively teach diverse groupings of students who bring different starting points, understandings, and styles of learning to the learning enterprise (Darling-Hammond, 1989, 1991, 1994; Garcia and Pearson, 1993; Oakes, 1985).

Redesigning the State Assessment System

Over the last several years New York State has been developing a comprehensive system of goals, curriculum, instruction, and assessment that supports a more robust view of learning than uniform content coverage for all students. This system is aimed at alleviating the tensions between teaching and assessing that have been created by testing systems of the past. It aims to support "bottom up" changes at the classroom level that both document and encourage students to understand and use knowledge in complex, meaningful, purposeful settings. At the same time it is designed to address "top-down" needs for reliable public state-wide information about student progress. Toward this end the State has developed rigorous standards for learning that are articulated in seven different curriculum frameworks. A new assessment system is currently under development to reflect and measure these standards. Building on New York State's long tradition of teacher involvement in the development and scoring of assessments, teams of teachers and assessment experts are designing examinations in a variety of subject areas that are comprised of two components: an on-demand test (time-bound examination in which students answer a uniform set of short, structured tasks, open-ended questions, and/or essays) and a common extended task that is contextualized and embedded in the curriculum (a project, portfolio, or performance that is worked on by students over time during the course of the school year). Each of these components are constructed to directly measure dimensions of the standards. The information about learning that these exams provide will inform the public about how students and schools are progressing toward attainment of the standards. In addition to these state examinations, local districts will be required to develop their own set of complimentary assessments (performance tasks and portfolio collections of student work) that are to be used primarily to inform instruction and support student learning (New York State Curriculum and Assessment Council, 1994).

When the State began this redesign process in 1992, it was faced with the difficult challenge of how to make change, how to bridge new practices with old. Innovative teachers in the field cried out for some kind of relief from existing policies and pressures so that they could genuinely change the ways that they teach. Their calls for change underscored the reality that *what is tested determines what is taught.* Changes in teaching cannot occur if they are not supported by changes in assessment; and likewise, changes in assessment cannot work effectively if they are not supported by changes in teaching.

One strategy developed to jump-start the change process was an "Invitation to Invention" - The Regents Exam Options Project. Initiated as a transitional policy to support change, it provided high school teachers and schools engaged in local curriculum and assessment reform efforts with temporary relief from some of the powerful and constraining aspects of the existing, highly-structured Regents system. The idea was that while the State redesigned its curriculum and assessment system, teachers (who were already practicing many of the performance-based teaching and assessment strategies that were being endorsed by the State) were invited to use these to count toward a percentage of their students' Regents examination credit. In schools that accepted this invitation, anywhere from 10-35 percent of the Regents grade in any particular discipline could be assigned to projects, performances, and portfolios embedded within the context of daily classroom learning.

Particular attention in this redesign process was paid to the State's high school Regent's examinations because as these exams are currently configured they are dominated by multiple-choice and short-answer questions that focus almost exclusively on content coverage. Innovative teachers and schools throughout the state have consistently complained that the demands of the Regents examinations have been a barrier to curriculum development and the use of teaching strategies that emphasize critical thinking and creative problem-solving. Teachers cite the following problems with the exams:

> Coverage demands of the Regents examinations make it difficult for teachers to spend the time that is needed to explore different topics in depth and to ensure that their students understand and can use the knowledge they are learning.
>
> The multiple choice format of these tests limits students to uniform responses often disengaged from meaningful, purposeful contexts that relate knowledge and information to real-life situations. This format also provides little opportunity to demonstrate such higher-order skills as analysis and synthesis, or the ability to make judgments wisely.

The press to prepare students for these tests often keeps teachers from engaging their students in creative studies and projects that engage them in the real-life tasks that are needed to successfully enhance their knowledge and skill acquisition.

Local Options as a Strategy for Change

The Options Project began in the 1993-94 school year with 19 projects in 10 school districts. Because it allowed a variety of "untested" innovations to count as credit towards a high-stakes assessment, it immediately encountered considerable criticism about its ability to ensure that student performance across the State would be measured in comparable ways. This was a legitimate concern since there was no way to ensure at this point that the Projects could indeed provide such data. But that was not the Project's original intent. It was initiated to address a different issue - to lessen the constraints that the Regents examinations imposed on teaching and learning in the classroom. As long as the Regents exam held high stakes for students, teachers argued, they could not take risks to depart from the fast-paced, highly-structured Regents curriculum in order to experiment with new and different ways of teaching. Offering credit for innovation in a percentage of the exam was meant to alleviate a little of this pressure. At the same time, it was hoped, the Project could be used to build teachers' capacities and understandings about building and scoring performance tasks - information that would be helpful in the redesign of the future State assessment system as well as the design of local assessments. It was also reasoned that since, at most, the Project would account for as little as 10-35 percent of the exam, there would still be enough uniformity in the rest of the exam to have some basis to compare students' performance.

Meanwhile, efforts were made to build some uniformity into the design of the Options by requiring each project to assess students according to the State syllabus or, in cases where they were completed, the new State standards for the discipline. State Education Department staff provided teachers with professional development supports to enable them to meet these requirements.

The Regents Exam Options Project was always intended to be phased out as the new State examination system emerged. Local variations to support better teaching would no longer be as pressing a need when a new, more flexible system of performance-based assessments would be available to support a greater range of students to demonstrate State standards in more varied ways.

Purpose of this Study

This study was designed to learn about the impact of the Options Project: the effect it has had on teaching and learning as well as its consequences as a strategy for implementing change. Did the permission to innovate that the Project granted actually promote better teaching and learning for students? Did the lack of comparability in the assessment design have any harmful effects on students or schools? What other changes occurred as a result of the Project's presence in schools?

To answer our questions we looked at the Options Project's first year. We asked the districts involved to complete reports that answered a detailed set of questions. We also conducted interviews, extensively sampled student work, and surveyed all the Project's teachers, students, and administrators.

What we found is that the Project has had a powerful impact on the schools that were involved. It has been a catalyst for changes in pedagogy, student and professional learning, as well as changes in the structures and culture of schools. It has developed a cadre of teachers who are both knowledgeable about and skillful in performance-based assessment development. Many of these teachers are now participating in and providing leadership to the State-led development teams currently engaged in designing new Regents pilots exams. The valuable understandings that they have gained from their own local experience is directly contributing to the development of a system of Regents examinations featuring common performance-based tests and tasks for all students across the State. At the same time, the Project has impacted positively on the learning and achievement of a broad range of learners.

Our study has led us to encounter significant issues and important questions about performance-based assessment systems. Can there be large-scale assessment systems that do not standardize all aspects of the assessments? How much variation in content, context, format, and procedures can be tolerated to be able to report comparable data about student learning while still maintaining authenticity in the assessment situations? Can teacher judgment be involved in decisions that determine high-stakes consequences for students? What kinds of assessments are fair for all students? What kinds of assessments best support the purposes of learner-centered schooling? These issues will be discussed in the body of this paper.

Project Descriptions

The Mousetrap Racing Car

A Physics class in a central New York district spent three weeks on a Regents Options Project to design a racing car using a mousetrap. This project was to count for

20% of the Regents examination score. Groups of students were to design a car that utilized essential physics principles - potential energy, kinetic energy, and gravitational potential. When completed, the cars were placed in competitions to see which won for highest speed, longest distance, and best climbing up a hill. Specific criteria for the design were made public to all students. Each student was required to produce a portfolio of drawings, labs, and performance data to complement his or her group's car building effort. Grades were awarded to individual students not on the basis of the car's performance but on the data that each one collected and the individual analyses that were made. The portfolio included:

1)Design information - scored for 6 points (Can it work? On what principle? Identification and evaluation of the potential problems with modification recommendations);

2)Prototype information - scored for 6 points (about construction, testing, problems identified, solutions proposed);

3)Construction and testing of the final car - scored for 0 points so as to emphasis each student's analysis and understanding of what happened);

4)Measurement and competition reports - scored for 8 points (reporting on potential energy stored, kinetic energy of moving car, gravitational potential energy in the hill climb).

In a survey of students who were part of this project, many had positive things to say. They commented on the project's ability to offer opportunities for collaboration with classmates. They also liked the fact that the project asked them to utilize their knowledge of the disciplines by solving problems in real-life contexts:

> This project was a lot of fun. It gave us the opportunity to work "hands-on" instead of just taking notes. It gave us the opportunity to learn how to work with our classmates to solve a problem. It kind of pumped your brain thinking of all the solutions to all the problems we had with our car. It is kind of rewarding to know you can solve a problem and build something using only materials and your brain, not specific directions.

> The project made us work together in a group and help each other when we needed to solve problems. We had to think a lot because no one was there telling us what to do all the time. We had to be responsible and mature adults. This was good because it gave us a chance to use our imagination and creativity.

> I like the idea of seeing physics in real life rather than taking a test. This was like a review of what we had learned about Potential and Kinetic Energy. Now I understand these even more than I did when they were

taught to me. At first it seemed like the labs were difficult, but once we got into them they were easy to understand and conduct.

The mousetrap car projects taught many of us to respect engineering and the laws of physics. It taught us that in designing something you must have 1)a plan of action 2)the drive and materials to make it happen 3)back up plans in case of malfunctions.

The best way to judge if someone understands what is taught in the classroom is if they can incorporate it into a real life situation. This worked well to test our physics knowledge and allowed us to see what we learned in the classroom and how to use outside it also.

Teachers too had positive reactions about the project. They noted how it provided a variety of opportunities for students to demonstrate their knowledge of and understandings about physics:

(From a science teacher): I observed several types of energy transfer setups. The cars showed that the students had considered many variables in their designs. For example: Holes were drilled in the car body to cut down mass; wheels were made very narrow to reduce friction; the powered axle had large diameter wheels so that one turn of that axle would move the car a greater distance.

(From a math teacher): I witnessed students working in groups of three or four refining their designs to accomplish different tasks: the car that could travel the longest distance, the fastest car, and the one that could best climb a hill. I hear comments from the students such as "Needs more resistance! Put rubber cement, on the wheel so the car won't slip so much." It was nice to see all of the students working on some aspect of their project without having to be coerced by their teacher. I also witnessed the groups giving each other suggestions to help solve a problem that one group was experiencing.

The Doomsday Curriculum

An Earth Science teacher in an outlying suburban New York City district developed a "Doomsday Curriculum" for her 35 percent Regents Option. Students were presented with the following dilemma: The earth was destroyed, making it necessary to find another planet on which to live. Each student was required to study a planet in depth for several months and to prepare a presentation about how to survive on it. Working in teams as well as individually, they used a variety of resources and materials and formats to pursue their learning. A final exhibition at the end of the year presented the students' different survival planets. Each team had to defend its work.

The teacher who developed this Option recollects with pride on how preparation for the project's final exhibitions created a culture of collaboration in her classroom, alleviating the highly competitive atmosphere that usually prevails in Regents and AP classes:

> They stood up together - from the brightest to the resource student - and what I saw was kids speaking and sharing with their peers with a sense of maturity, ownership and pride.

The Language Arts Portfolio

Regents Language Arts classes in the 11th grade of a New York City public high school designed a portfolio that was counted as 35 percent of the Regents grade. It consisted of four pieces of writing: a response to poetry; a critical analysis of prose or drama; an informative essay responding to some 'real world" situation; and a reflective letter in which the student discussed his/her writing style, strengths and weaknesses. These four pieces were designed to replace 50 multiple choice questions of the current Regents exam: the vocabulary, spelling, and reading comprehension sections. The assistant principal, who led the portfolio effort, explains how the writing assignments were constructed to be more challenging than the old multiple choice test:

> They were modeled on the AP exam. They certainly required more work and they put the students under more pressure. Despite this, an analysis of the results shows us that the 290 Juniors with portfolios scored higher than Seniors from the previous year who did not compile portfolios. We are especially happy with the scores earned by students in PCEN [remedial] classes.

Changing the Ways that Teachers Teach

100 per cent of the schools involved in the Options Project reported that their participation improved learning for students and impacted positively on teaching. They cited a variety of ways that the development and use of performance assessments impacted on learning.

The Options Projects helped teachers to transform their roles from transmitter of information to facilitator of learning. By providing students with opportunities to choose their own formats for demonstrating their learning and engaging students in group processes rather than only individual work, teachers were enabled to break away from a teacher-centered mode of instruction to a more student-centered one. One teacher commented on this transformation:

> This project changed my role as a teacher. It allowed a different interplay between and among students and teacher. Our roles were constantly reversing. I learned as much from what they came up with as I taught. I became a coach. Because the kids were so invested in their projects there were times when I could have walked out of the room, gone to lunch, and then come back and they'd still be there working. This freed me to get up, walk around, and listen to conversations. If I felt some students needed extra support, I could ask a question and keep them moving. I also could dialogue with my bright kids who were zooming off in all kinds of directions with "what ifs, what ifs."

Other changes noted in classrooms as a result of the Projects were greater use of cooperative learning strategies;

> Cooperative learning flourished and students received lots of practice working together.

increased emphasis on problem-solving and higher order thinking skills;

> The pedagogy is entirely different. We spend more time learning how to write a persuasive essay, how to think critically, how to reach for that higher standard for all the kids.
>
> As the project got under way, I found a sudden shift in what I was teaching. I was now referring more to the *laws* of physics rather than simply presenting mathematical equations.

more opportunities for students to engage in active, inquiry-oriented learning;

> I have focused my attention more on the importance of discovery and performance activities in science rather than on "cookbook" style learning.
>
> I have emphasized hands-on, problem-solving, discovery - oriented activities most suited to the students in my class. Encouraging creativity in the ways in which students can demonstrate knowledge is an important objective.

a broader range of ways for students to illustrate mastery of skills and understanding of content (including the use of art, music, videos, and dramatic presentations); and more occasions made available for students to engage in work that had personal relevance and meaning. As one teacher explained:

> If students' choice of topics can be used as a measuring stick for the Global Studies Authentic Assessment Option for my students, the relevancy of this evaluation tool can easily be seen. The most common topics were rain forest depletion, Somalia, and HIV-related topics. Many

students, however, from high ability to low, chose to pick the "path less taken." Unique topics were researched and presented, such as street children of Brazil, women's rights in Japan, robotics in Japan, and the disaster in Chernobyl and its aftermath. The key to all, whether rain forest or robotics was that the students perceived the relevancy, i.e., making history a living and contemporary topic. From that standpoint and how it related to the essay writing on the Global Studies Regents, I would judge the authentic assessment model developed by Global Studies teachers to be a very successful teaching-learning tool.

These assessments, like others that call upon students to perform in ways that are much like those needed in real-world situations, challenge teachers to break down the traditional barriers that have been erected between real-life and school-based learning. To prepare students to perform well on these assessments, teachers must teach not only information, but how that information can be used. They must build the capacities of their students to work with others, to explain their ideas, to question assumptions, to create new ways of solving problems. These are the skills that students will need as they enter the workplace and the future. These are the skills that were emphasized in the Options Project performance-based assessments.

Supporting Student Achievement at Higher Levels of Performance

The Options Project impacted positively on student achievement as well as on students' attitudes towards their learning. Both teachers and students confirmed this in interviews. The overwhelming majority of them thought the Options Project was a worthwhile experience, with 95% of the students in one urban district in central New York State reporting that they felt this way. The small percentage of students in this district who were not in favor of the projects were described by their teachers as those who traditionally attained high test scores with seemingly little effort or work. These students seemed to resent the increased demands and more rigorous standards that the Options Project required of them.

Despite the changes in the overall Regents exams brought about by the Options Projects, student scores remained constant in comparison to past years. Pass rates in all the projects did not vary from previous years by more than a few percentiles. What *did* change however, was the distribution of the scores. While the number of students scoring above the 50th percentile remained about the same, the number of students scoring above the 25th percentile increased substantially (for example, 44% in 1993: 83% in 1994). What accounted for these changes? One reason offered by many students and teachers was that the performance task directly assessed how students were able to use their knowledge instead of indirectly testing knowledge through the

fill-in-the-blanks responses to questions that traditionally appear on tests. One science course for example, used its Option to have students engage in *actual* laboratory experiments instead of completing the multiple choice questions *about* laboratory skills that were on the traditional exam. For this performance task 95% of the students scored in the top third percentile as opposed to 65% of the students who scored in the top third for the multiple choice questions in the previous year. Is it any wonder that students were more successful in *demonstrating* lab skills in real performances rather than answering questions *about* lab skills on a reading/writing test?

Even in the districts where adoption of the Options Project resulted in lower examination scores, administrators and teachers insisted that the increased challenges to students presented by the performance assessments were definitely worth the decline in scores.

> The Options Project forces the student who really should be doing more challenging work to do it.

Other changes in student learning outcomes were noted by districts participating in the Options Project. A number of districts noted that students became more self-directed and took more responsibility for their learning. 94 percent of the teachers in one large urban district reported that the assessment project was noticeably empowering to the students.

> They are more aware of the expectations and can work on the assessment independently.

This phenomenon was attributed to the assessments' clearly defined and publicly articulated standards for what students were expected to know and be able to do. These provided students with a structured focus that enabled them to be clear of what was expected of them. As one teacher stated:

> Students were actively trying to achieve on an assignment that was "doable" in carefully supervised, incremental steps, with constant feedback provided.

Even the librarians in the Options Project schools reported almost unanimously that they too witnessed changes in student life - students appeared to be more challenged, and to have increased motivation for their work. The librarian in one large district noted that, in spite of the greater difficulty level of the Project than other work

students normally received, students not only showed increased motivation but also seemed to enjoy it more. In hypothesizing about why this was so, she noted the increased challenges that the projects presented, especially to traditionally low-achieving students:

> I think the biggest change for kids is that they spend a lot more time figuring out how to do things and not as much time learning *pieces* of information. They thought that the last four weeks doing an interdisciplinary kind of problems approach was much more fun than trying to memorize pieces of information so they could get the usual ninety-five on the Regents. We've gone for the high ground in designing our exam, for using the AP as the model, for saying we're going to give all of our kids a greater intellectual challenge. We're gonna bring this challenge all along the line. The people that have gained the most from this are probably our below-average kids that now have a chance to try a lot of these things that they've been pushed off from before. None of them failed to succeed in this.

Student involvement in what was considered by all to be more challenging work resulted in greater student self-esteem and increased respect for each other. One district found this especially evident as students prepared for their final assessment exhibitions:

> Students displayed pride as they presented. They also demonstrated curiosity regarding their own topics and those of their classmates.
>
> Students responded very positively to presentations of their fellow classmates, listening attentively and reacting positively. The Projects reflected very favorable results - the best essays of the year.

Students also learned to use their talents and to share their talents with others. In developing their portfolios and class presentations, many students broke out of the competitive culture that often dominates high track classes. They critiqued each other's work and coached each other in much the same way that adults do in collaborative settings. This helped to create a culture of collegiality in which all were moving toward excellence together.

> Everybody in the room would be working on a project and what would happen is that someone would say, "Gee I'm having a problem with my topic" and then somebody would overhear and say, "Oh, I know exactly what you need."

This culture of collegiality was infectious, spilling over to affect interactions among teachers as well. As one teacher said:

> This kind of thing is not only happening now among kids but is also happening among teachers.

The power of learning through meaningful work - that is the outcome most frequently mentioned by all the participants in the Options Project. The Project provided students with occasions to investigate topics that had direct relevance to their interests and circumstances. In one biology class, for example, a student who is a diabetic did his report on diabetes; another wrote about back problems since he is afflicted with difficulties himself. Through these different entry points, students learned many of the important principles of biology. Even though the content differed, the same important standards of the discipline were stressed and demonstrated in the projects. Allowing students to show their understandings of biology through work that incorporated their choice of subject, empowered students to take ownership of their learning. The teacher of this class concluded this in an interview. He said:

> The students owned their own assessment.

In the same interview he went on to describe how the investment in learning that grew out of students working on issues that were especially meaningful to them led several students involved in a joint plant growth project to continue the project's lab on their own even though they did not intend to submit it for credit.

> By all accounts what we did, what kids did, and what teachers did was very powerful. There is something that is absolutely fascinating about having a whole class doing some very meaningful things.

Enhancing Teachers' Professional Learning

Articulating values and goals for schooling and developing assessments that demonstrated if and how students were able to meet those goals was a learning experience for teachers. Engaging in this process helped them to clarify and develop their thinking about teaching, about student learning, and about curriculum development. Together they built shared norms and standards for their collective professional practice.

Designing tasks and scoring rubrics based on standards challenged teachers to closely examine their own teaching. They questioned if they had been teaching what is important; if there were focus and cohesiveness to their program. Pursuing these questions challenged them to be able to more clearly articulate both the purpose and

the strategies of their pedagogy. As a result, it heightened teachers' standards for their own performance.

Designing performance assessments also informed and enhanced teachers' pedagogy. Assessing what students knew and could actually do in meaningful and purposeful contexts helped teachers look closely at their teaching strategies, at curriculum, at the learning environments they constructed, and most importantly, at students themselves. Many began to look at students and their work in broader ways so that they came to know students better, to see and understand the "varied array of mental competencies, strengths, or "intelligences' that they possess" (Kornhaber and Gardner, 1993, p.2-3). This led them toward better understanding of learners and learning and away from reliance on uniform prescriptions for teaching. They learned more about effective teaching strategies, expanded their teaching repertoires, and extended their reach to successfully engage a wider range of students to achieve more challenging standards.

Making standards clear and publicly articulated also raised teacher expectations of students which, in turn, helped students to raise their own sights. The publicly-articulated criteria for learning helped students to know what constituted excellence and what they needed to do to attain it. This was facilitated in many cases by displaying samples of student work that demonstrated different levels of performance and the diverse possibilities for what could be considered excellent. And because the assessment projects were constructed to legitimize multiple ways of knowing and to acknowledge diversity and the contextual nature of learning, they created possibilities for students themselves to have greater access to achievement and success.

Work on development of the Options Project also increased teacher interaction across disciplines and departments, breaking down the isolation that has been a traditional part of teaching. Teachers could not do this project without collaborating, dialoguing, and reflecting with other colleagues. As they talked with each other about their students, their teaching strategies, their curriculum, and the learning environments they were constructing they found themselves developing new ideas, new ways in which to implement their ideas, and new structures that would support their practice. One large upstate city school reported:

> The amount of enthusiasm, rich dialogue and reflective teaching that was occurring in classrooms around the city was enormously gratifying. Teachers everywhere were stepping back, rethinking, retooling.

discarding the old, inventing the new and rediscovering the joys of teaching.

Other districts reported similar dynamics:

> This project allowed me to recognize the value of working with teachers of other disciplines toward common products of student work and hopefully this will lead students to make better connections among the subjects they learn from teachers of different disciplines.
>
> For the first time that I can remember, teachers in different departments are working together and talking about instruction.
>
> Teacher cooperation was stressed in all respects, setting rules, grading procedures, sharing work loads, sharing lesson plans, discussing texts and assignments. This was widely felt to be an important outcome.
>
> We had an opportunity to work with colleagues on a common project. It was very rich. The fact that we all just really brainstormed, worked very hard to come up with a plan was a collegial professional activity that doesn't happen all the time and that's one of the real benefits that comes out of this.
>
> We had unexpected ties and unexpecting comings together and I think that's very empowering. We were continually helping each other.

The overwhelming majority of teachers who participated in this project reported it to be an experience that was professionally and personally transformative. It challenged them to rethink their roles, their actions, and the ideas that have driven their practice.

> If you've taught for a while you know you can be very successful at what you're doing. But are you willing to kind of drop down into this sort of new teacher role again and be able to put up with some of the failures that occur? Have you become somewhat complacent with how you teach and would like a stimulating experience that would give you a fresh look at teaching and learning? Well this is gonna be right up your alley. It will turn you upside down, it will change your life regardless of what happens to your kids and what happens in your classroom. It will make you an entirely different teacher, hopefully for the better.

Change leads to more change

Exposure to new forms of assessment through the Options Project has heightened teachers' understanding of the interconnection between teaching, curriculum and assessment. As one teacher said:

> This is only a beginning. The ideal and ultimate goal will be to remove the separation of testing and learning as much as possible so that we may be able to assess better what students really know. Then students will begin to see learning as having value in itself and not in the artificial number or letter grade which has replaced learning and has itself become the goal.

The process of working on the Options Project has also led many teachers and administrators to see the need for more comprehensive changes in schools aimed at better supporting student learning. Some teachers have become so convinced of the benefits of performance-based assessments that they are participating in State-led efforts to redesign the entire Regents examination system; some schools have introduced performance assessment strategies to teachers and students in other disciplines; some teachers have begun to design comprehensive performance-based assessment systems for their classes that include many types of demonstrations of student learning - performance tasks, portfolios, and projects as well as the traditional kinds of tests; some schools have begun to experiment with other changes in school life that facilitate the kind of real-life learning promoted through performance assessments - interdisciplinary teaching, changes in scheduling and other uses of time, team structures, collaborative teaching, and parent/community involvement.

One district moved toward interdisciplinary work creating a new team structure of ninth grade teachers (science, English, math, and social studies) to assist children, particularly those with academic and social needs, in making a successful transition from the middle school to high school. Several other districts restructured their schools into block scheduling plans so that students and teachers could have more time to engage in the in-depth learning that the performance assessment projects required.

Other districts increased their efforts to educate and involve members of their community about new curriculum and assessment practices. The changes in school practices that resulted from the move to performance assessments made them realize the importance of communicating with all members of their school community. They instituted Exhibition Nights for parents and the community to share the student work developed through the projects and also made presentations to their Boards of Education. As an administrator in one of the large New York State cities said:

> If you start off with an idea that you're really gonna go to performance assessment, that changes the way you teach, it changes the way you operate with kids. You need to do more student -student dialogue, teacher-student

dialogue, dialogue with everybody concerned and it has to go on and it has to go on continuously.

Challenges for Performance-Based Assessment Systems

It seems clear from the evidence presented here, supported by other studies of schools that have used performance assessments, that performance assessments which are embedded within the contexts of everyday learning have the capacity to provide rich and broad information about student learning that has educational benefits for teachers and students (Darling-Hammond, Ancess, and Falk, 1995; McDonald, et al., 1993; Mitchell, 1992; Perrone, 1991; Shepard, 1995; Valencia, Hiebert, and Afflerbach, 1994; Wolf, 1989; Wood and Einbender, 1995). Performance assessments encourage curriculum and teaching that support students in developing high-quality, individualized products and performances. They offer diverse opportunities for sustained and deep engagement in learning. By legitimizing multiple ways of knowing, acknowledging diversity, and utilizing the contextual nature of learning, they provide students with access to more valid and varied means by which to demonstrate their abilities.

The data from this study as well as other assessment research suggests however that a variety of issues, problems, and challenges must be addressed in order for performance-based assessment *systems* to be viable.

One of the biggest challenges has to do with attitudes and prior conceptions of learning. Some students involved in assessment projects were uncomfortable taking charge of their own learning, having grown accustomed to traditional patterns of student-teacher interaction in which students passively receive instruction on material determined by the teacher. Such students need extraordinary support to learn how to motivate themselves and to think on their own. Still other students, particularly those who generally have done well through the use of traditional teaching and assessment, were not happy with the fact that they were required to do more work on the assessment projects than on traditional assessments in order to maintain comparably good grades. Because in many New York State schools the Regents exam grade determines if students pass a course, students with a bent toward traditional "test-taking" have been able, in the past, to pass the Regents test (and thus pass the course) while actually doing little work for the class. Schools involved in the Options Project required students to perform beyond the test. This change in expectations toward more rigorous standards means that students need to clearly understand just what it means to be academically successful.

In spite of this fact, most students reported that even though it was easier to take a "regular" multiple-choice exam, they liked the performance assessment better. Their responses to survey questions indicated that the performance assessment made school more interesting, more challenging, and provided them with a greater range of opportunities to demonstrate their learnings.

Time will most probably resolve these tensions as students get used to learning and being assessed in new ways. Time is always needed to help people adjust to any change. It is also needed to address other issues. Many teachers and administrators expressed the need for more time to be provided within the school day for students to engage in the in-depth work that made up the substance of the Options Projects. Students need extended periods of time to engage in research, inquiry studies, and laboratory experiments. The traditional 42 minute period does not allow for engagement in such serious learning. Teachers too need time to prepare, learn, and collaborate with others as they develop and work through projects. They also need time to assess the work - to review and reflect on their students' products, performances, and processes, to become familiar with scoring procedures, and to actually evaluate the end results. Reporting the results in ways that help to inform the community about the process also requires additional time.

A variety of other supports are needed by teachers if they are going to be able to simultaneously develop new kinds of assessments *and* make changes in their teaching. They need flexible scheduling, learning materials and resources to conduct the projects, and opportunities for collaboration with their colleagues. They also need to be supported to take the risks required to engage in change. We note that administrators, in planning for these supports, will need to consider that many teachers are in different places in the growth continuum - ranging from being eager to make changes to being reluctant - and that plans will need to be developed keeping these differences in mind. Without such strategic planning it will be difficult to ensure that students have had the opportunities to learn what the new assessments are measuring.

Tensions Between Reliability and Validity

In addition to issues of implementation and change there are a whole set of important technical issues that must be addressed to enact performance-based assessment systems. Despite the growing evidence about the consequential validity of these assessments, there is still great reluctance to use them in large-scale testing systems. Opponents of the use of contextualized performance assessments in

In spite of this fact, most students reported that even though it was easier to take a "regular" multiple-choice exam, they liked the performance assessment better. Their responses to survey questions indicated that the performance assessment made school more interesting, more challenging, and provided them with a greater range of opportunities to demonstrate their learnings.

Time will most probably resolve these tensions as students get used to learning and being assessed in new ways. Time is always needed to help people adjust to any change. It is also needed to address other issues. Many teachers and administrators expressed the need for more time to be provided within the school day for students to engage in the in-depth work that made up the substance of the Options Projects. Students need extended periods of time to engage in research, inquiry studies, and laboratory experiments. The traditional 42 minute period does not allow for engagement in such serious learning. Teachers too need time to prepare, learn, and collaborate with others as they develop and work through projects. They also need time to assess the work - to review and reflect on their students' products, performances, and processes, to become familiar with scoring procedures, and to actually evaluate the end results. Reporting the results in ways that help to inform the community about the process also requires additional time.

A variety of other supports are needed by teachers if they are going to be able to simultaneously develop new kinds of assessments *and* make changes in their teaching. They need flexible scheduling, learning materials and resources to conduct the projects, and opportunities for collaboration with their colleagues. They also need to be supported to take the risks required to engage in change. We note that administrators, in planning for these supports, will need to consider that many teachers are in different places in the growth continuum - ranging from being eager to make changes to being reluctant - and that plans will need to be developed keeping these differences in mind. Without such strategic planning it will be difficult to ensure that students have had the opportunities to learn what the new assessments are measuring.

Tensions Between Reliability and Validity

In addition to issues of implementation and change there are a whole set of important technical issues that must be addressed to enact performance-based assessment systems. Despite the growing evidence about the consequential validity of these assessments, there is still great reluctance to use them in large-scale testing systems. Opponents of the use of contextualized performance assessments in

accountability systems argue that they lack sufficient standardization to determine if students in different locations have reached comparable levels of achievement. They further contend that the range of interpretation and response is so great in complex performance tasks that it is impossible to generalize across groups of students to ascertain if they have achieved, with any consistency, similar bodies of knowledge and collections of skills (Moss, 1994).

Traditional testing has avoided this problem by standardizing the format, the content, the context, and the procedures for the completion of exam problems or tasks. By relying heavily on the use of multiple choice or right/wrong answer formats, it has reduced ambiguity about student performance. To make scoring easy, unbiased, and uncontroversable, test questions have been simplified of complexity and context.

Constructing tests in this manner however has exacted heavy tools on teaching and learning. Because the tests measure only what is easily measurable, instruction has been similarly reduced. Teaching in many schools and classrooms becomes little more than preparing students to perform well on the severely constricted format and content of the tests (Wiggens, 1993).

Traditional attempts to standardize tests place a host of other limitations on the ability to assess learning, resulting in assessments that acknowledge only a narrow spectrum of performance, sacrificing the individuality of the learner, the process of learning, the nuances that genuinely reflect what and how learning is achieved. This kind of assessment does not provide students with vehicles to demonstrate the different kinds of competencies or intelligences that they may possess. It disadvantages some students over others, denying those with divergent strengths the possibility of demonstrating excellence in their own ways (Darling-Hammond, 1994; Darling-Hammond, Ancess and Falk, 1995; Kornhaber and Gardner, 1993; Falk, 1994; Falk, MacMurdy and Darling-Hammond, 1995; Garcia and Pearson, 1994).

In contrast, authentic information about student learning is provided when students experience worthy questions and tasks that faithfully represent the diverse and rich contexts encountered in any particular field of study. Authentic information about student learning comes from examining performance in real-life situations - situations that call upon students to use a wide repertoire of skills, to justify answers as well as solve problems, and to create quality products which meet clear and demystified standards for work.

Such representation of reality in assessment, while positively impacting teaching and learning, does indeed, create technical difficulties if the information is to be used for accountability purposes. Authentic assessments *are* difficult to compare. It

is hard to generalize across groupings when assessment responses are complex and take place in varied contexts. Yet, if this tension is not addressed, student performance will continue to be evaluated in ways that do not reflect the full range of capabilities and understandings of diverse learners. We will continue, in the name of psychometric reliability, to sacrifice authentic and valid learning.

But how do we resolve the inescapable tensions between reliability and validity that performance assessments pose for assessment systems? *Can* assessments that are individualized, genuine analogs to real-life learning also be used as indicators of progress in large-scale settings for high stakes purposes? What is the capacity for such assessments to be reliable and generalizable across classrooms, schools districts, and states? Do we have to exchange reliability for validity or can there be some way for a system to address them both?

We argue, as others have already done (Moss, 1994; Wiggens, 1989, 1993). that traditional measures of student learning have heretofore achieved psychometric reliability at the expense of validity. It is not good enough, we contend. to be able to compare student achievement across groupings if the achievement is only measured by the limited indicators of traditional standardized tests. Assessments used for accountability purposes *must* include meaningful indicators of student learning. These are obtained through performances that are embedded in life-like situations and that allow students flexibility in determining how they demonstrate what they know and can do (Glaser and Silver. 1994; National Forum on Assessment. 1995). If we sacrifice this in assessment design for the sake of yielding unambiguous results that can be easily compared, we privilege those who are able to display their competencies in the singular format determined by the testmakers, disadvantaging those who are divergent thinkers or whose strengths and abilities are best expressed in different ways. If we choose reliability over validity we turn our backs on diversity, thus perpetuating inequitable access to high standards and an inequitable distribution among our population of those who are able to achieve excellence.

Finding Authentic *and* Comparable Measures of Performance

The Regents Exam Options Project demonstrates how context-embedded performance-based assessments are able to reflect the knowledge and understanding of diverse learners about important standards for learning. The challenge it presents for the design of assessment systems is to find a way to use these assessments to also compare student performances among and across groups.

We suggest that this challenge be addressed by creating an assessment system that at once balances uniformity with authenticity and context. We believe that this balance can be achieved by creating performance assessments based on uniform standards in the disciplines. These assessments can evaluate student achievement of important dimensions of learning through common scoring rubrics that articulate different levels of student performance. The assessment tasks themselves should be required to embody some common elements regarding format and procedures, but their content, context, and outcomes should vary to reflect differences in how individuals learn and express what they know. While "there is no such thing as an authentic task" (personal conversation with Robert Linn, June 28, 1995), careful attention to ensuring that even when students are all asked the same questions they are always provided with opportunities to express what they know in their own way. can help to safeguard the ability of assessments to reflect the highly individualized nature of understanding.

Scoring of student work should be done by teachers, those closest to the learning situation. Moderation exercises can be conducted at district and regional levels across a state to ensure that teachers' judgment is a reliable and valid means to measure student progress. Early trials of moderation studies with multi-faceted assessments in California and New York State, suggest that teacher judgment is indeed a reliable way to assess student attainment of standards (Barr, 1995; Klausner, 1995).

New York State's assessment system design features tests that do exactly this. The "on-demand" component is to be administered to students under similar conditions, calling on them to respond to uniform questions but also providing them with opportunities to construct and explain their own answers. The "extended task" component is embedded in the classroom curriculum. It requires common elements (i.e., all students must do a research project or science experiment that contains the same three parts), but also allows room for variation to ensure that the assessments are responsive to individual differences in contexts, interests, and learning modalities. Preliminary findings from the first round of pilot exams indicate that teacher agreement on scoring has a high degree of consistency.

An assessment system designed in this way establishes a basis for generalizing about student performance and achievement within and across large and diverse groupings. It offers opportunities for students to demonstrate what they know and can do in both uniform and contextualized settings, maintaining an element of standardization without sacrificing authenticity. It relies on teachers to assess student

performance, a process that values their professional judgement while at the same supports their professional learning.

Data from the Options Project points to the power of these kinds of assessments for supporting teaching and learning in schools. More work needs to be done in the future to ensure that such assessments adequately measure intended standards, are designed in ways that allow for student performance to be accounted for consistently and comparably over time, and yield scores that represent reliable indicators of student performance.

As understandings continue to increase about the kind of education that supports a greater range of citizens to achieve at higher levels of performance, all educators are challenged to also expand our definition of what constitutes sound assessment. We need to develop assessments that do not constrain and constrict teaching and learning. We need to develop assessments that support what we have come to know is needed for learning: opportunities to use and apply knowledge, to inquire, to analyze, to critically evaluate, and to use creativity to pose and solve problems. We need to make these assessments an integral part of the learning experience, allowing students to demonstrate in a variety of ways, suitable to their individual talents, what they really know and can do. Only when real and meaningful student work is made a part of the assessment process can there be valid and equitable evaluation of the skills and abilities of all students.

References

Barr, M. (1995). Who's going to interpret performance standards? A case for teacher judgment.. Paper submitted to *Claremont Reading Conference 59th Yearbook*. Claremont, Ca: Institute for Developmental Studies of the Claremont Graduate School.

Darling-Hammond, L. (1989). Curiouser and curiouser: Alice in testingland. *Rethinking Schools, 3*(2), 1, 17.

Darling-Hammond, L. (1991). The implications of testing policy for educational quality and equality, *Phi Delta Kappan. 73*, 3, 220-225.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review, 54* (1), 5-30.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action*. New York: Teachers College Press.

Darling-Hammond, L., Einbender, L., Frelow, F., & Ley-King, J. (1993). *Authentic assessment in practice: A collection of portfolios, performance tasks, exhibitions, and documentation*. New York: NCREST.

Falk, B. (Ed.) (1994). *The Educational Forum*, 59 (1) - Issue on Authentic Assessment.

Falk, B., MacMurdy, S., & Darling-Hammond, L. (1995). *Taking a Different Look: How the Primary Language Record Supports Teaching for Diverse Learners*. New York: National Center for Restructuring Education, Schools, and Teaching.

Garcia, G.E. & Pearson, P. D. (1994). "Assessment and Diversity." In L. Darling-Hammond (Ed.), *Review of Research in Education, 20*, pp. 337-391. Washington, D.C.: American Educational Research Association.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Glaser, R. & Silver, E. (1994). "Assessment, testing and instruction: Retrospect and prospect." In L. Darling-Hammond (Ed.), *Review of Research in Education, 20*, pp. 393-419. Washington, D.C.: American Educational Research Association.

Klausner, E. (1995). The Primary Language Record as a reliable option for assessing children's literacy: Report of the initial phase of development of a process for moderation. New York: The New York Assessment Network.

Kornhaber, M and Gardner, H. (1993). *Varieties of excellence: Identifying and assessing children's talents*. New York: NCREST.

McDonald, J., Smith, S., Turner, D., Finney, M., & Barton, E. (1993). *Graduation by exhibition*. Alexandra, VA: Association for Supervision and Curriculum Development.

Mitchell, R. (1992). *Testing for learning: How new approaches to evaluation can improve American schools*. New York: The Free Press.

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher 23*(2), 5-12.

National Forum on Assessment. (1995). *Principles and indicators for student assessment systems*. Cambridge, Mass.: Fairtest.

New York State Curriculum and Assessment Council. (1994) *Learning-Centered Curriculum and Assessment for New York State*. Albany, New York: New York State Education Department

Perrone, V. (Ed.) (1991). *Expanding student assessment*. Alexandria, VA: ASCD.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. Yale University Press: New Haven.

Resnick, L. B. (1989). *Tests as standards of achievement in schools.* Paper presented at the Educational Testing Service Conference, The Uses of Standardized Tests in American Education, New York.

Shepard, L. (February, 1995). "Using assessment to improve learning." *Educational Leadership*, pp. 38-43.

Valencia, S., Hiebert, E., and Afflerbach, P. (1994). *Authentic reading assessment: Practices and possibilities.* Newark, Delaware: International Reading Association.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 70*, 703713.

Wiggens, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing.* San Francisco: Jossey-Bass.

Wolf, D.P., (April 1989). Portfolio assessment: Sampling student work. *Educational Leadership*, pp. 35-39.

Wood, D. & Einbender, L. (1995). *An authentic journey: Emergent understandings about authentic assessment and authentic practice.* New York: NCREST.

Participants in the New York State Regents Exam Options Project 1994

Auburn High School

Bethlehem Central Senior High School

Buffalo School District

Croton-Harmon High School

Guilderland High School

Humanities High School, New York City

Marlboro High School

Mineola High School

Ossining High School

Plainedge High School