DOCUMENT RESUME

ED 424 293 TM 029 179

AUTHOR Snow-Renner, Ravay
TITLE Mathematics Assessment Practices in Colorado Classrooms: Implications about Variations in Capacity and Students' Opportunities To Learn.
PUB DATE 1998-04-00
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Information Analyses (070) -- Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; Educational Practices; Educational Research; *Elementary School Teachers; Elementary Secondary Education; *Mathematics Education; *Secondary School Teachers; Surveys; Tables (Data); *Teacher Attitudes
IDENTIFIERS Authentic Assessment; *Colorado; *Opportunity to Learn; Reform Efforts

ABSTRACT

To measure the extent of student opportunities to learn relative to educational reform goals, the Colorado Educational Policy Consortium at the University of Colorado-Denver designed and administered teacher and student surveys about instructional and assessment-related processes statewide. This study uses data from the 1997 teacher survey to explore teacher reports about assessment practices in mathematics classrooms relative to student opportunities to learn. Survey responses were received from 737 mathematics and science teachers in grades 4, 8, and 10 (approximately 17% of the teachers for those grades), and 116 elementary school and 223 secondary school students provided information about their classroom assessment practices. The emphasis of elementary school teachers on authentic assessments was greater than that reported by secondary school teachers. However, overall findings indicate that students in different classrooms experience differential opportunities to learn relative to reform-oriented assessments, and that teachers indicate varying levels of capacity for implementing such assessment practices. Such variation may be partially attributable to fluctuations in teacher capacity and knowledge and partially to ambiguous policy definitions of reform in Colorado. Implications for further study include extending research about student opportunity to learn through alternative examinations of classroom-level assessment practices, and greater investment in building local teacher capacity. (Contains 6 tables and 54 references.) (SLD)

Mathematics Assessment Practices in Colorado Classrooms: Implications about Variations in Capacity and Students' Opportunities to Learn

Ravay Snow-Renner
Colorado Educational Policy Consortium
University of Colorado, Denver

Recent standards-related reform movements in mathematics and science call for sweeping shifts in the nature of classroom instruction and assessment, and have driven educational policy in many states. In Colorado, such policies include Colorado's House Bill 93-1313, which officially has adopted "standards-based education" for the state and for individual school districts, and the Colorado Student Assessment Program, initiated in the spring of 1997, which involves statewide testing of almost all students at given grade levels. However, linked with the more commonly-understood reform components of *content* and *performance standards* is the element of *delivery standards*, or the assurance that all students enjoy an equitable *opportunity to learn* the materials upon which they are being measured. Student opportunity to learn is a broad concept, highly dependent upon classroom interactions, and operationalized not only in terms of content coverage, but relative to student exposure to complex and demanding modes of assessment.

In order to measure the extent of student opportunities to learn relative to reform goals, the Colorado Educational Policy Consortium (CEPC) at the University of Colorado-Denver was contracted to design and administer teacher and student surveys about instructional and assessment-related processes statewide. This study uses data from the 1997 teacher survey to explore teacher reports about assessment practices in mathematics classrooms relative to student opportunities to learn. Findings indicate that students in different classrooms experience differential opportunities to learn relative to reform-oriented assessments, and that teachers indicate varying levels of capacity for implementing such assessment practices. Such variation may be attributable partially to fluctuations in teacher capacity and knowledge, and partially to ambiguous policy definitions of reform goals in Colorado. Implications for further study include extending research about student opportunity to learn, partially through alternate examinations of classroom level assessment practices, and also greater investment in building local teacher capacity, through providing teachers themselves with more opportunities to learn how to use these complex assessment tools.

An introduction to standards-based reforms and implications for assessment

Standards-based education, as it has been conceptualized in policy and research documents (National Council on Education Standards and Testing [NCEST], 1992; Conference Report, 1994; McLaughlin & Shepard, 1995) entails several different categories of standards. The most highly publicized have been *content standards* and *performance standards;* however, another critical component of standards-based reform is *delivery or opportunity-to learn standards.*

Content standards

Content standards are broad depictions of the skills and knowledge that students should acquire and be able to do in a given subject area, and are perhaps the most publicly understood aspect of the reforms (McLaughlin & Shepard, 1995). Following the lead of national mathematics and science education groups, such as the American Association for the Advancement of Science, the National Council of Teachers of Mathematics, and the National Research Council, states, districts, and even individual schools have created standards writing teams in different subject areas to develop general statements about what their students should know and be able to do at different levels. In Colorado, state-level writing teams composed of content experts, educators, and community members developed a series of draft documents in each content area. Each draft was made available to the public for input or approval, and, based upon responses, revised for another iteration of the process. The final State Model Content Standards were then approved by the state School Board and avowed as the quality benchmark that local school districts needed to "meet or exceed" with their own, locally-drafted standards documents.

Concomitant with the implementation of content standards in the classroom, although not explicitly stated in Colorado policy documents, these reforms also call for changes in instruction based upon constructivist ideas about learning, ideas that involve more cooperative student grouping structures and more active learning in classrooms. In mathematics, documents used as models for the content standards development process, such as the National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards for Mathematics* (1989) and *Professional Standards for Teaching Mathematics* (1991) call for a shift in mathematics from a curriculum emphasizing computation and rote memorization of facts and procedures to one that is conceptually oriented, engaging all students in developing mathematical power. Under this vision, students are engaged in construction of knowledge through conjecture, analysis, and application of mathematics in real-world and mathematical contexts.

In 1992, the science education community began to convene groups of science educators and scientists to develop standards for science curriculum, teaching, and assessment under the aegis of the National Research Council. Building upon earlier works such as AAAS' Project 2061 and the National Science Teachers Association's Scope, Sequence, and Coordination Project, the group developed the *National Science Education Standards.* This document expresses a vision consistent with that of the *NCTM Standards*. Both the *NCTM Standards* and the *National Science Education Standards* agree that science and mathematics education should:

- Emphasize high expectations for all students;
- Engage students in meaningful activities that enable them to construct and apply their knowledge of key science and mathematics concepts;
- Reflect sound principles from research on how students learn, including the use of cooperative learning techniques promoting interaction and deeper understanding;
- Feature appropriate, on-going use of calculators, computers, and other technologies for learning science and mathematics;
- Ensure that teachers have a deep understanding of their subject matter; and
- Provide ongoing support for classroom teachers, including continuing opportunities for teachers to work with one another in planning curriculum, instruction, and assessment (Weiss, 1994).

Performance standards

Performance standards may be characterized as more specific examples and explicit definitions of knowledge and tasks that students must successfully complete in order to demonstrate mastery of the content standards. These standards are typically exemplified through the nature of the assessments used to measure student achievement. When one considers the breadth of change in instruction implied by the reforms, it is clear that the implications for related changes in assessment practices are equally sweeping. Additionally, the technical issues are likely more formidable, especially when standards-related assessments may be used for accountability or certification purposes.

Since content standards are to exemplify complex, higher-order skills and thought processes, it is argued that using the same sorts of low-level, multiple-choice standardized assessments that have been historically used for ranking and measuring students over the years is inappropriate. Alternative assessments are needed. Over the past ten years, researchers have put forth a variety of proposals for alternative assessment systems, based upon reform emphases on "higher-order" thinking skills (or ambitious content standards) and research showing the corruptive effects that widely-used standardized, multiple-choice assessment measures have on such ambitious learning goals.

Resnick and Resnick (1992) influenced early conceptions of standards reform by suggesting that complex assessments should be used to drive improvements in instruction. They reviewed the historical relationship between assessment and instructional programs, and concluded that current multiple-choice standardized achievement tests of basic skills drive curriculum and instruction toward low-level expectations of students. As an alternative, they advocate using performance assessments to measure higher-order thinking and content which will drive instruction toward what they call the "thinking curriculum." Such types of assessments, which might encompass the

use of rubrics, portfolios, or student-generated projects, would be considered more "authentic" (or integrated within classroom instructional practice) than more standardized measures (Wiggins, 1989; Shepard, 1989). They also are theorized to be potentially more "systemically valid" (Frederiksen & Collins, 1989) in that they are less easily corruptible than high stakes standardized measures; improved student test scores are more likely to validly represent student learning, rather than other score polluting factors induced by a high stakes testing environment (Haladyna, et al, 1991).

These recommendations for changes in assessment all emphasize the importance of cultivating an educational and social environment where individuals have the capacity to recognize that there are different purposes and technical requirements for assessments, and where the use of multiple, complex measures to determine student achievement are encouraged. Ideally, new measures would be embedded within the classroom curriculum, rather than imposed externally, and would not interfere with the course of higher-level instruction in the negative ways that norm-referenced high-stakes assessments have been shown to do, especially in classrooms with high proportions of minority students (Baron, 1990; Rottenberg & Smith, 1990; Lomax et al, 1992).

The new forms of assessment that are being developed in response to these critiques of standardized assessment practices imply a broad transformation of the conceptualization of test validity and the relationship between testing and instruction. Resnick and Resnick describe performance assessments as " tied to the curriculum and *designed to be taught-to.*"(p. 72). This characterization contradicts common practice and beliefs about validity and the relationship between norm-referenced, standardized tests and instructional programs.

Teaching to the test has traditionally been viewed as cheating because it violates assumptions underlying norm-referenced, standardized test item construction. Norm-referenced test items represent a domain of content that is generalized across different curriculum and instructional treatments. Students who are taught or prepared to respond to specific test items violate the assumption that a test item is only a sample of the knowledge domain, and their scores therefore do not accurately reflect learning across the entire domain. Norm-referenced test items are not constructed to be taught-to and are not valid for comparing students across instructional programs if some students are taught the items. However, new assessments, such as performance assessments, which are designed to measure the quality of a complex synthesis of important skills that comprise a knowledge domain, actually require being "taught-to" in order to be valid and fair. These fresh interpretations of validity and fairness are integral to a conception of delivery, or opportunity to learn standards.

Opportunity to learn

Opportunity to learn (OTL), or *delivery standards*, in addition to ambitious content standards and performance standards that measure student achievement, comprise the last integral element of standards-related reforms. According to the National Academy of Education's report on standards-based education (McLaughlin & Shepard, 1995), opportunity to learn standards:

> define the level and availability of programs, staff and other resources sufficient to enable all students to meet challenging content and performance standards. "Opportunity" comprises such things as teachers who are well prepared in their content area, instructional materials and resources adequate to instructional goals, a safe school environment, and courses and instructional activities consistent with more demanding standards of content and performance (42).

OTL provides new ways of thinking about equity and due process in relation to standards-based assessments. The criterion for equity in standards-based education goes beyond the traditional definition of equity as equal resources, measured as either equal spending per pupil or equal taxable resources, to more specifically address whether resources are *adequate* to enable students to learn what is expected and assessed (Clune, 1995; Smith and O'Day, 1990; NCEST, 1992). Standards-based assessments are not considered valid or fair if students have not had adequate opportunities to learn what they are expected to know and be able to do.

Federal legislation in Goals 2000 calls on states to develop criteria for judging whether all students have adequate opportunities to learn what they are expected to know and be able to do in the standards. These criteria should assess:

> the sufficiency of quality of the resources, practices, and conditions necessary at each level of the education system (schools, local educational agencies, and states) to provide all students with an opportunity to learn the material in voluntary national content standards (Conference Report, 1994).

This policy definition is abstract and ambiguous because it was negotiated through consensus among policy makers with competing and conflicting ideas about standards, assessments and OTL. The result is bipartisan political support for the reform in the abstract, leaving states with broad discretion for defining and using standards, assessments, and OTL. Consequently, definitions of OTL vary widely by locality. Additionally, the requirements for what constitutes "adequate" opportunity to learn in assessments vary depending upon the intended use of assessment results.

Uses of tests differ in different conceptions of standards-based education. Resnick and Resnick (1992) indicate that assessments and results should drive instruction, and OTL describes the extent to which students are provided appropriate and adequate classroom instruction. However, other standards proposals emphasize using test results for accountability of systems, teachers or students (Smith and O'Day, 1990; Shanker, 1994). Drawing upon Messick's (1989) conception of *consequential validity*, Linn (1994) explains psychometric principles of validity in performance assessments and standards-based assessments as a judgment about the uses and interpretations of the results rather than about a test. As test results are used for accountability or certification purposes, technical requirements that OTL is adequate become more stringent. Particularly if high stakes are attached to individuals' test results, the focus changes "from 'What students know and can do' to 'What students know and can do as a result of their educational experiences'" (Burstein and Winters, 1994, cited in Muthen, Huang, Jo, Khoo, Goff, Novak and

Shih, 1995, p. 371), and data about those educational experiences are required in order to estimate the validity of the measure.

Operationalizing OTL across these shifting contexts is a complex task. Most OTL data are gathered through teacher surveys because they are cost-effective and easily administered in conjunction with student assessments. In general the OTL that has been examined most frequently through national and international teacher surveys include content/topics covered and the format and context of the content covered. Frequently, researchers in assessment have expressed concerns over OTL issues relative to access to the thinking curriculum that would prepare students to do well on such complex, progressive assessments as those described by Resnick and Wiggins (Herman, 1997; Darling-Hammond, 1995; Madaus, 1991; Herman, et al, 1996; Smith, 1994; Winfield & Woodard, 1994). Dennie Palmer Wolf, in treating assessment as a "learning event," makes explicit the link between OTL and the very mode of student assessment itself, as well as content coverage (Wolf, 1993; Wolf & Reardon, 1996). Wolf and Reardon assert that only by providing universal access to such meaningful, higher level assessments can we ensure that students will have equal opportunities to learn, and that teachers develop a common language to define performance and shape instructional strategies, as alternatives to the simplistic teaching and assessment practices supported by the use of standardized or otherwise externally-imposed tests. Using this broader definition, OTL as it applies to standards-based assessments may be construed to include not only sufficient exposure to the *content* tested, but also *exposure to the testing format* (e.g., open-ended answers, narrative explanations about reasoning, estimation and speculation).

Standards-related reform in Colorado

In 1993, Colorado enacted legislation establishing "standards-based education" statewide, HB93-1313. The bill was drafted using the rhetoric of support for the types of higher-order classroom interactions (evaluation, synthesis of ideas) characterized by the Resnicks as "the thinking curriculum," while also specifying the nature of content to be taught in mathematics, science, and other core content areas. State-level teams were organized to draft and revise state model content standards in six different First Tier areas (mathematics, science, reading, writing, history, and geography), with drafts subject to public input and review. Once the model standards were finalized, each of Colorado's 176 school districts was required to create and approve its own set of local content standards in the same areas, which were to "meet or exceed" the state standards in quality, or to adopt the state standards outright. A sample of Colorado's fourth standard in mathematics is provided on the following page.

As Colorado is a strong local-control state, and representatives of local districts are specifically granted discretion over instructional practices and textbook selection (Colorado Constitution, Article IX, Sections 15 and 16), the changes in instruction and assessment advocated by national groups such as the NCTM and the NRC were not explicitly included in the legislation.

STANDARD 4:

Students use geometric concepts, properties, and relationships in problem-solving situations and communicate the reasoning used in solving these problems.

In order to meet this standard, a student will

- connect various physical objects with their geometric representation,
- connect mathematical concepts from across the standards with their geometric representations;
- recognize, draw, describe, and analyze geometric shapes in one, two, and three dimensions;
- make, investigate, and test conjectures about geometric ideas; and
- solve problems and model real-world situations using geometric concepts

(Colorado Department of Education, 1995)

In order to measure student performance relative to these new goals, and partially approved as an accountability measure, the Colorado State Assessment Program (CSAP) was introduced into law in 1997. Initial student assessments took place in the spring of that year in reading and writing at fourth grade and most students in the state at grade level were tested. The assessments, while largely multiple-choice, incorporate more constructed-response and open-ended items than in the past, and were designed to measure more complex, higher-order processes than traditional multiple-choice measures. State-level assessments in other content areas, among them, mathematics and science, are scheduled for upcoming years. Although the state has recommended that the results of the CSAP not be high stakes for students, results are available at the individual student level, and thus hold the potential for being used in ways for which they are invalid. Data addressing student OTL relative to the CSAP measure have not been collected, so it is impossible to estimate validity of the measure for high-stakes purposes.

Especially when one considers Wolf and Reardon's (1996) characterization of the mode of assessment as part of a student's opportunity to learn, the necessity for examining current assessment practices across the state appears vital. This study appraises mathematics teacher reports about classroom assessment practices and examines the implications for students' opportunities to learn accordingly. What are mathematics teachers doing in Colorado classrooms, in terms of their assessment practices? How may this relate to student OTL? What are existing levels of teacher capacity for implementation and what may be needed to improve classroom assessment practices?

Instrument design and sampling procedures

Because of the statewide scope of inquiry, a survey was chosen as the primary measure to examine teacher assessment practices. In 1993, the Colorado Educational Policy Consortium (CEPC) began the design of a comprehensive survey for mathematics and science teachers and students in Colorado. Individual items were derived from national and international surveys addressing constructivist reforms in mathematics and science, including the National Survey of Science and Mathematics Education (Weiss, 1993), the Schools and Staffing Survey (NCES, 1993), a Stanford-based survey of elementary mathematics teachers in California (Center for Research on the Context of Teaching, 1994), NAEP, the National Assessment of Educational Progress (ETS, 1992), and the Survey of Mathematics and Science Opportunity, administered in conjunction with TIMSS, the Third International Mathematics and Science Study (International Association for Evaluation of Educational Achievement [IEA], 1994).

Using a process similar to that described by Blank (1993), the measure's scope was gradually refined and modified to account for needs specific to the context of Colorado's own reform. For instance, response formats for assessment items were adjusted from frequency reports (1-3 times per month, for example) to percentages of total time spent on assessment, based upon responses to the pilot items. Additionally, items addressing content coverage in terms of student opportunities to learn were re-tooled so that they mapped specifically onto the state mathematics and science content standards. After two pilot administrations, the surveys were revised for baseline data collection use in May, 1996. The instruments were administered again in April and May of 1997. Sampled groups were Colorado mathematics and science teachers at grades 4, 8, and 10, the same grade levels at which the CSAP had originally planned the state testing. This study uses data from the 1997 administration of the Colorado Teacher Survey.

A stratified random sampling strategy was devised that targeted all Colorado secondary schools and a random sample of elementary schools. Surveys were distributed by building principals and teacher respondents were provided with anonymous, postage-prepaid, preaddressed envelopes so that they could mail completed survey materials directly to the CEPC. To maximize response rate, no identifying codes were used on survey materials, although teacher data and student data were linked. 737 teachers (approximately 17% of all teachers in the targeted population) responded statewide, and participants appear reasonably representative of the state as a whole. (The findings reported here, however, should not be taken to generalize beyond the three populations sampled--4th, 8th, and 10th grade mathematics teachers). 339 respondents (116 elementary teachers and 223 secondary teachers) provided information relative to their mathematics assessment practices in the classroom.

Assessment practices in the classroom

In Colorado, as throughout the rest of the United States (NCES, 1996), mathematics teachers are providing students with a mixed bag of learning opportunities relative to reform recommendations and practice. They report using a variety of instructional strategies, similar to the "melange" of traditional and reform-oriented practices that David Cohen described in his case

study of Mrs. O (Cohen, 1990). There are some significant differences of pedagogical practice by instructional level; for instance, elementary teachers report significantly less time on lecturing in mathematics than secondary teachers, and more time in student use of manipulatives, similar to findings elsewhere. According to both elementary and secondary teachers, however, between 9 and 11% of their instructional time over a semester is spent on testing, as defined in the traditional sense of testing (classroom tests and standardized tests).

Reports about specific assessment practices

Of this time spent on assessment activities, elementary and secondary teachers were asked to describe the proportions of assignments or tests that could be described in certain ways (e.g., tests that are performance-based, tests that use memorized rules and formulas). Averages and standard deviations are shown below.

Proportions of mathematics assignments or tests that...
(totals can exceed 100%)

	Elementary Teachers		*Secondary Teachers*	
	M	*SD*	*M*	*SD*
have more than one answer or approach	30.74	26.91	37.09	28.78
require students to apply what they have learned to real life situations or problems	40.10	27.95	33.43	24.39
require students to apply concepts or principles they have learned to new situations or problems	28.34	23.75	27.26	23.76
are performance based	43.27	30.23	43.28	65.55
are evaluated with a rubric	28.04	31.99	19.61	26.39
require students to provide a narrative explaining their reasoning	21.58	22.86	16.07	21.69
require students to explain their reasoning orally	30.55	26.36	17.07	21.01
demonstrate basic skills/vocabulary	44.30	29.24	32.94	29.85
use memorized formulas and rules	30.93	26.22	26.66	25.73
require students to evaluate and improve their own work	38.27	30.03	33.11	30.74
require student to conduct investigations over several days	17.67	22.30	11.87	16.69
become part of a portfolio of students' work	21.86	29.57	21.23	33.51

Table 1

Teacher reports demonstrate a mixture of pedagogical practices. Both elementary and secondary teachers report fairly high proportions of assessments that measure basic skills and vocabulary, and slightly less use of assessments using memorized formulas and rules. They also report considerable emphasis upon more progressive assessments that have more than one approach, require students to evaluate and improve their own work, require application of knowledge to real-life or different problems, require students to evaluate or revise their own work, or are performance-based. Additionally, elementary teachers report considerably more emphasis on oral explanations of student reasoning and the use of rubric-evaluated assessments than secondary teachers. To examine differences by instructional level more closely, a series of ANOVAs were run, with the following significant results:

- Secondary math teachers report more use of tests with more than one answer or approach ($F = 6.672$, $df = 1, 445$; $p = .010$)

- Elementary teachers report more tests that require application of knowledge to real life ($F = 5.348$, $df = 1, 453$, $p = .021$) and more requiring oral explanations ($F = 34.954$, $df = 1, 427$, $p = .000$) and narrative explanations ($F = 5.816$, $df = 1, 433$, $p = .016$) of student reasoning. They report more use of reform-oriented assessments, such as rubric-evaluated measures ($F = 5.716$, $df = 1, 415$, $p = .017$) and measures that take several days to complete ($F = 7.993$, $df = 1, 403$, $p = .005$). However, and perhaps as might be expected, they also report significantly more emphasis on tests demonstrating basic math skills ($F = 11.15$, $df = 1, 434$, $p = .001$).

At first blush, these findings appear plausible, although some results are incongruous. For example, the reporting of performance-based assessments does not appear to function as it was intended, reflecting the comprehensive definition of performance-based assessments as measures in which rubrics for performance are designed and used in professional development to enhance generalizeability across scorers. Were this the case, reports about performance-based assessments and rubric-evaluated measures should be more congruent. Reports about "performance-based assessments" average roughly 43%--a relatively high figure--for both elementary and secondary teachers, compared to approximately a 19% to 24% average on assessments that are evaluated using a rubric. It seems likely that semantic issues are at play in the general prompt about performance-based assessments, as will be discussed below.

The relatively large amount of variance in responses pointed up the need to examine the data in more complex ways than simply by comparing means. Response frequencies were examined and organized into five groups; teachers who reported that none of their classroom assessments fell under that category, and then teachers whose responses fell into quartile ranges (signifying less than one-quarter, 26-50%, 51-75%, or more than 75% of the assessments or assignments used in class are of the pertinent type). Table 2 on the following page shows these results; frequencies for all mathematics teachers are shown unless preliminary ANOVA's showed significant differences in teacher responses. For these variables, frequencies are displayed by instructional level.

Type of mathematics assignment or assessment...	*Percentage of All Teachers Reporting*				
	Not used	*1-25%*	*26-50%*	*51-75%*	*76% +*
require students to apply concepts or principles they have learned to new situations or problems	1.3%	61.9%	24.7%	5.6%	6.5%
are performance based	4.4%	39.9%	24.8%	9.1%	21.8%
use memorized formulas and rules	5.4%	57.6%	20.5%	9.5%	7%
require students to evaluate and improve their own work	6.4%	46.2%	24.6%	8%	14.5%
become part of a portfolio of students' work	35.9%	38.4%	11.3%	1.3%	13.1%

Type of mathematics assignment or assessment...	*Percentage of Teachers Reporting by Level*				
	Not used	*1-25%*	*26-50%*	*51-75%*	*76% +*
have more than one answer or approach					
Elementary teachers	5.3%	51.9%	25%	7.9%	9.9%
Secondary teachers	2.4%	49.1%	24.1%	9.5%	14.9%
require students to apply what they have learned to real life situations or problems					
Elementary teachers	.7%	42.4%	31.4%	13.1%	12.4%
Secondary teachers	.3%	52.3%	29.2%	11.9%	6.3%
are evaluated with a rubric					
Elementary teachers	28.2%	37.3%	14.1%	4.9%	15.6%
Secondary teachers	25.1%	53.1%	10.5%	3.3%	8%
require students to provide a narrative explaining their reasoning					
Elementary teachers	11%	63%	16.4%	5.2%	4.1%
Secondary teachers	11.8%	72.6%	8.3%	2.5%	4.8%
require students to explain their reasoning orally					
Elementary teachers	5.3%	52.3%	25.2%	7.9%	9.3%
Secondary teachers	14.7%	67%	10.8%	2.2%	4.3%
demonstrate basic skills/vocabulary					
Elementary teachers	1.4%	37.6%	29.5%	15.1%	16.4%
Secondary teachers	3.1%	55.5%	20%	7.3%	14.1%
require student to conduct investigations over several days					
Elementary teachers	21.2%	59.8%	11.7%	3.7%	3.6%
Secondary teachers	19.4%	70.5%	6.7%	1.2%	2.2%

Table 2

Variation in assessment practice becomes apparent when one examines the columns showing extreme levels of use or nonuse (the particular assessment practice was not used at all or was used more than 75% of the time). Almost 36% of math teachers across levels report that no classroom assessments or assignments have become part of a portfolio of student work over the previous semester. Fourteen percent of secondary teachers and more than 16% of elementary teachers report that most (more than three-quarters) of the classroom assessments that they have used over the semester measure basic skills and vocabulary. More than one-quarter of mathematics teachers at elementary and secondary levels report that they have not used assessments that are evaluated using rubrics at all over the previous semester, although almost 22% report using

assessments or assignments that are "performance-based" most of the time. Additionally, a substantial proportion (almost 15%) reported that students were required to evaluate and improve their own work more than three-quarters of the time, a primary theme of assessment reformers.

Exploring teacher reports through factor analysis

To further examine the ways in which these variables functioned, a factor analysis was conducted. Primarily the analysis was confirmatory, to test initial hunches about "traditional" assessment practices (e.g., basic skills assessments, uses of memorized formulas and rules) and more "reform-oriented" practices (e.g., performance-based assessments, rubric-evaluated assessments, portfolio elements, multi-day investigations); however, it was also designed to shed light upon how certain items were functioning. All assessment variables tended to intercorrelate significantly (of all bivariate correlations among the 12 variables, only five were insignificant) and significant correlations ranged in size from .110 to .636. Due to these high intercorrelations, factors were computed using the Principal Components analysis (criteria for factor selection = eigenvalue > 1) and an oblique rotation. Preliminary analyses indicated that factor structures were similar across instructional levels, therefore, all cases (both elementary and secondary) in which teachers reported about mathematics assessment practices were included in the analysis.

Three factors emerged, accounting for 56.178% of total variance. The first factor, *Authentic assessment practices*, corresponded closely to the recommendations of assessment reformers, including rubric-evaluated assessments, requirements that students provide narrative or oral explanations of their reasoning as part of the assessment, portfolio assessments, and student investigations that last over a period of several days. The second factor, *Applied and complex assessment practices,* included practices that required students to apply their knowledge to new or different situations, practices with more than one answer or approach, and assessments that were "performance based." Two of the three variables loading on the third factor were clearly *Traditional assessment practices*--focusing on basic skills and memorized formulae, but the third was more problematic, as it involved student evaluation and revision of work. Table 3 shows factor structure, variable loadings, and subscale reliability on the following page.

Apparent semantic issues with several items may have implications for estimating the extent of teacher capacity. The "performance-based" assessment item was based upon research about rubric use in performance-based assessment, and had been projected to load on the same factor (factor I) as the rubric item. However, it clearly functioned differently than expected, loading (.616) on factor II. Reliability estimates confirmed that this variable was behaving oddly; an analysis of the factor II subscale indicates a reliability of .5466 (Cronbach's alpha). When the "performance-based" variable is omitted from the scale, reliability goes up to .7245.

When examined in conjunction with frequencies reported (almost 22% of teachers reported that more than three-quarters of their classroom assessments were "performance based," although only approximately 8% to 16% said so many of their assessments were evaluated using rubrics), these data indicate that the "performance-based" variable is likely being interpreted in its broadest sense. This may plausibly include an indication of paper and pencil "performance", or filling in the

correct multiple-choice option, rather than "performance" in terms of complex accomplishments similar to those elicited by New Standards Project tasks. Validity studies of teacher surveys on opportunity to learn, conducted by Leigh Burstein, et al (1995) have found that teachers do not have common understandings of assessment terms, especially when items are broadly phrased, as this item is. It is likely that the more specific "rubrics" item provides more valid information about teacher practice relative to "performance-based" assessments in the sense of the reforms. However, the way in which this item is functioning raises questions about the extent to which teachers understand assessment reform recommendations, and hence their level of capacity to implement such strategies.

Mathematics Assessment Practices

	Loading	*Scale Reliability*
Factor I--Authentic Assessment Practices		.7434
Students are required to provide a narrative explaining their reasoning	.760	
Students are required to conduct investigations over several days	.727	
Students are required to explain their reasoning orally	.684	
Assignments or tests become part of a portfolio of students' work	.658	
Assignments or tests are evaluated with a rubric	.650	
Factor II--Applied and Complex Assessment Practices		.5466
Students are required to apply what they have learned to new situations	.790	
Students are required to apply what they have learned to real life situations or problems	.760	
Assignments or tests have more than one answer or approach	.639	
Assignments or tests are performance-based	.616	
Factor III--Traditional Assessment Practices		.6669
Assignments or tests demonstrate basic skills and vocabulary	.807	
Assignments or tests use memorized formulas and rules	.784	
Students are required to evaluate and improve their own work	.709	

Table 3

Similar semantic issues arise with the item addressing student evaluation and improvement of their own work. This is a major theme of assessment reformers, involves student recognition and ownership of criteria determining quality, is based upon constructivist theory, and arises in discussions of literacy (Hiebert & Raphael, 1996), portfolios and authentic assessments (Wiggins, 1989), and also in mathematics (Voigt, 1995). However, as with the "performance-based" variable, this item is functioning differently than intended. It is loading (fairly strongly--.709)with other variables that are clearly traditional and low-level, emphasizing memorization and basic skills. Additionally, it does not detract from the reliability (Cronbach's alpha = .6669) of the factor subscale. One plausible explanation for this may be that a teacher could respond that many classroom assessments have this characteristic because in class, his or her students, after a test or quiz, are frequently asked to "exchange papers and grade your neighbor's"--a fairly traditional timesaver for teachers. Although this is not the meaning intended by the prompt, it is possible

that it can be interpreted in this way, and the fairly high (14.5%) proportion of mathematics teachers saying more than 75% of their assessments fall into this category may support this hypothesis. Again, teacher understanding about reform issues appears unclear, due to the way this item is functioning.

Examining teacher reports in terms of capacity to implement assessment reform

Teacher reports about practice show variety in assessment practices around the state, supporting the hypothesis that differential OTL (in terms of access to demanding, complex mathematics assessments as part of the learning environment) is experienced by different students at grades 4, 8, and 10 across Colorado. At the elementary level, students of approximately 28% of these teachers never have the opportunity to work on math assessments that are evaluated using rubrics, although more than 15% of elementary teachers report that rubric-evaluated assessments are used at least 75% of the time in their classrooms. At the secondary level, rubric use is even less pronounced; only 8% of teachers report using rubric-evaluated assessments more than 75% of the time, and one-quarter of all secondary math teachers say that such assessments are never used in their classrooms. At both elementary and secondary levels, approximately one in five teachers reports that students are never required to conduct investigations that last several days for math class. More than one-third of all math teachers report that portfolio assessments are not used in their classrooms, although approximately 13% of their peers report that their students work on portfolio-oriented assessments frequently. The variability of these results certainly has implications for student OTL and variations in local and individual capacity for providing it in assessment practice across levels.

The data suggest that issues of vertical articulation need to be addressed, as well. In terms of activities that relate to *authentic assessment practices* (as operationalized by variables loading on Factor I), elementary teachers are significantly more progressive than secondary teachers. Subscale scores for factor I were generated, and a one-way ANOVA run to check on potential differences in assessment practice by level. As might be expected from previous results on individual variables, elementary teachers report higher proportions of assessment activities dedicated to more authentic and progressive practices ($F = 18.43$, $df = 1, 451$, $p = .000$). These findings are potentially validated by achievement results in studies such as the Third International Mathematics and Science Study (TIMSS). The TIMSS measure was designed to reflect reform recommendations about complex, performance-oriented assessments, and achievement results show a steady downward trend in U.S. mathematics achievement (measured normatively against other countries) as the student test-taking population advanced in age (TIMSS International Study Center, 1996; TIMSS International Study Center, 1997; Takahira, et al, 1998).

Additionally, the inconsistencies in teacher responses to several reform variables as addressed above indicate variations in interpretations of items. These potentially may indicate related variations in practitioner capacity to implement progressive assessment practices in their classrooms, and consequently, variations in student OTL.

Research about reform capacity

Capacity for educational change (in terms of the core technology of the classroom) has been addressed in numerous ways. Some researchers have focused on overall organizational features, such as the creation of new structures to provide policy support and incentives for education personnel to implement general change (Elmore, 1996; Conley & Odden, 1995). Others have focused more on the interactions between organizational structures and practitioner beliefs and attitudes (Jennings & Spillane, 1996; Spillane, 1994; Spillane & Thompson, 1997; Spillane, 1998) in conceptualizing the nature of local capacity for implementation of new, sweeping educational reforms.

Spillane and Thompson (1997) argue that local capacity for implementing reform can be examined in terms of capital--*resource capital, human capital,* and *social capital,* and they focus on the last two in their study. *Human capital* is characterized as professional (teacher) commitment to reform, drive, content expertise, and ability to teach other professionals about needed changes. It is considered pivotal in developing *social capital* for capacity-building, which is described as norms of collegiality and collaboration, and active participation in professional networks. Spillane and Thompson suggest that, without taking the relative strengths or weaknesses in these capital areas into account, state or nationally-generated policy reforms like the standards will do little to increase implementation capacity or to equalize it among school sites. They forecast that sites with capacity may become even richer and that those without will continue to lack even minimal capacity for reform--which has serious implications for equitable student OTL.

Other research has examined local capacity specifically within the context of assessment reform. Pamela Aschbacher (1993) identifies a series of specific barriers to and facilitators for the implementation of innovative assessments within the classroom. Factors facilitating meaningful change were:

- *teacher commitment to reforms*, characterized as "purposeful passion". Obversely, a barrier was general reluctance to change practices;
- *collegiality*--being part of a group of learners. However, one barrier related to this involved a *lack of time* available for teachers to actually construct the meanings of alternative assessment practices for themselves and to become comfortable and proficient in their use;
- *sustained technical assistance* in both assessment issues and basic cognitive theory and its implications for instruction. A *lack of training and ongoing support* from experts was cited as a barrier to change; and
- *administrative support* for the changes.

Aschbacher's findings especially emphasized the unexpectedly large investment of time and resources (examples of assessments, portfolios, rubrics, etc.) in terms of improving teacher understandings of the reform.

A study by Prestine and McGreal into the implementation of assessment reforms in Essential Schools (1997) attribute the failure of such reforms to similar factors. They note a lack of knowledge about and understanding of authentic assessment, (roughly analogous to Spillane and Thompson's *human capital*), prevailing norms of privacy and teacher autonomy that supported conservatism in assessment practice (contrary to the issues of *collegiality* addressed by Aschbacher), issues of inadequate time (also related to Aschbacher's findings about time and development of professional knowledge), and a *fragmented approach*.

Findings from a research study in Arizona about the state Student Assessment Program provided additional information about factors that may contribute to improving local capacity for assessment change (Smith, et al., 1997). The study found that, while responses to the program varied across the state, responses coherent with the intent of the reform were centered in a few places where circumstances were auspicious, or which had innate implementation capacity. Several important characteristics identified as contributing to local capacity were *material and knowledge resources*, characterized both in terms of *financial* and *human capital*--materials to purchase necessary materials and training time, and technical support, as well as individuals with expertise in assessment, and *assumptive worlds*, or the patterns of beliefs that characterize a particular site. These "assumptive worlds" included beliefs about student capacities, a theme echoed in Jennings & Spillane's study of variations in the implementation of special education legislation in North Carolina (1996), and beliefs about pedagogy.

Exploring components of "capacity" and their relation to progressive assessment practices

Data about assessment practices from the Colorado teacher survey may be further examined in relation to several aspects of capacity. In terms of *human capital*, data relative to teacher commitment to standards, rated alignment of classroom teaching and classroom tests with the math content standards, and extent of professional development relative to standards and assessments were collected. Variables addressing *social capital's* aspects of capacity were also utilized; in addition to resource questions about student opportunities to learn and teachers having adequate resources to help students meet the standards, several items addressed administrative support. Table 4 provides descriptions of the variables, on the following page.

Logically, these variables should display a coherent relationship to teacher reports about progressive assessment practices, consistent with the research. It was hypothesized, for example, that teachers who experience more opportunities to learn about these assessment practices, for example, should report higher proportions of classroom assessment spent on progressive practices, such as using rubric-evaluated materials or conducting investigations over several days. In order to examine relationships, the scale score generated for variables loading on factor I, *Authentic assessment practices*, was used.

Variables	*Type and Range (Lower numbers in response options indicate more negative relation to standards)*
Human Capital	
The standards are important to me in planning my classes.	Likert scale: 1-5
How well does your classroom teaching currently align with the district mathematics content standards?	Likert scale: 1-3
How well do your classroom tests currently align with the district mathematics content standards?	Likert scale: 1-3
In the past three years of teaching, about how many days of professional development and courses for college credit have you completed in mathematics standards, curriculum, assessment?	Ratio-level constructed response
What percent of your total amount of professional development reported above was spent on assessment/performance assessment related to standards?	Ratio-level constructed response
Social/Resource Capital	
All students in my school have the opportunities they need to achieve the mathematics content standards.	Likert scale: 1-5
Teachers in my school have what they need to successfully implement the mathematics content standards in their classrooms.	Likert scale: 1-5
The principal supports teachers to implement the standards in our classrooms.	Likert scale: 1-5
The district administrators support policies and practices related to the standards	Likert scale: 1-5

Table 4

Because elementary teachers' classroom emphasis on authentic assessments was greater than that reported by secondary teachers, analyses were run separately by level. In general, teachers tended to agree that standards were important in planning their classes, although elementary teachers attributed more importance to them than secondary teachers. At both levels, teaching and testing practices within the classroom were reported as fairly well-aligned with standards, with averages of between 2.15 and 2.42 on a 3-point scale, with a 3 indicating full alignment. (By way of contrast, the alignment of district and standardized tests, which was also measured, was rated much lower, with averages of 1.79 and 1.82 for elementary and secondary teachers, respectively, on the same 3-point scale.) Secondary mathematics teachers reported considerably more days of professional development around math standards, curriculum, assessment, and instruction than elementary teachers (more than 9 days over the past three years, compared to approximately 5 ½ days), and all teachers reported that approximately one-quarter of their professional development had been spent on assessment issues related to the standards. In terms of social/resource capital issues, while teachers across levels tended to agree that administrators supported the implementation of standards, there was considerably less agreement about whether teachers had what they needed to help all students meet the standards, especially at the elementary level. Means and standard deviations are provided in the following table.

Descriptive data about indicators of local capacity by instructional level

Indicator and size of scale	*Elementary*		*Secondary*	
Human Capital	*M*	*SD*	*M*	*SD*
rating--personal importance of standards in planning the math class (1-5)	4.0	.84	3.83	.99
rating--extent to which classroom teaching is aligned with standards (1-3)	2.39	.53	2.42	.55
rating--extent to which classroom tests are aligned with standards (1-3)	2.15	.62	2.21	.64
days of professional development on math standards, assessments, etc.	5.40	6.36	9.26	14.7
extent to which professional development has emphasized assessment (%)	24.69	20.37	23.30	20.80
Social/Resource Capital				
rating--extent to which students have adequate OTL (5)	3.58	1.10	3.72	1.08
rating--extent to which teachers' needs are met (5)	2.91	1.26	3.10	1.22
rating--extent to which principal supports standards (5)	4.25	.89	4.23	.88
rating--extent to which district administrators support standards (5)	4.11	.89	4.04	.99

Table 5

To examine potential relationships between different variables and scale scores on the *authentic assessment* factor, bivariate correlations were run, again for elementary and secondary mathematics teachers. Table 6 illustrates these correlations and significant correlations are flagged.

Correlations between indicators of reform capacity and authentic assessment practices

	Correlations	
Indicator	*Elementary*	*Secondary*
Human Capital		
rating--personal importance of standards in planning the math class	.179*	.208**
rating--extent to which classroom teaching is aligned with standards	.067	.170**
rating--extent to which classroom tests are aligned with standards	.102	.165**
days of professional development on math standards, assessments, etc.	.222*	.334**
extent to which professional development has emphasized assessment	.095	.173**
Social/Resource Capital		
rating--extent to which students have adequate OTL	-.020	-.005
rating--extent to which teachers' needs are met	.081	-.005
rating--extent to which principal supports standards	.012	.112
rating--extent to which district administrators support standards	.030	.007

* Correlation is significant at p<.05
** Correlation is significant at p<.01

Table 6

As may have been expected, Aschbacher's "purposeful passion" (1993) appears to play a role in teacher assessment practices; higher ratings of the importance of standards in planning teachers' classes correlated significantly with more progressive assessment practices. Additionally, the extent and focus of professional development around mathematics standards and assessment practices experienced by teachers tended to correlate significantly with more progressive assessment practice, especially for secondary teachers. For secondary teachers as well, ratings of how well classroom teaching and testing practices were aligned with math content standards correlated significantly with more progressive assessment practice, but this was not the case for elementary math teachers.

However, all of these correlations, while significant, are fairly small in size. Even the strongest relationship, with an r^2 of .334 between days of professional development about math standards, assessments, and instructional practices and the use of authentic assessment practices by secondary teachers, is only small to moderate in size. In practical terms, these relationships are not much to write home about. Further explorations of the data, in terms of initial regressions using the variables in this model, indicate similar findings; although viable and statistically significant predictive models have been generated, the amounts of variance in assessment practices that they explain are negligible, ranging from 10.2% for elementary teachers to 12.6% for secondary teachers.

Possible explanations and implications for further study

There are a variety of explanations for these findings. Certainly measurement error may have contributed to and confounded them. However, given the presence of correlates that correspond to research about local capacity, it seems likely that these data have provided a fairly reasonable representation of what teachers think they are doing in the classroom. One probable explanation for the fairly small relationships between capacity indicators and progressive assessment practices is that teachers *do not share clear understandings about what the Colorado standards reform involves, especially in the area of related changes in assessment practices.* Such a conclusion is consistent with other research findings around assessment reform (Smith, et al, 1997, Herman, 1997); although most teachers report that they are conforming to the requirements of these reforms, they do not possess deep understandings of them.

This hypothesis is supported by a variety of evidence. While the unexpected functioning of the two reform-oriented variables addressing "performance-based' assessment and student-evaluated work cannot be interpreted as wholesale evidence that teachers do not understand these concepts within the context of assessment reform documents, it does demonstrate differential interpretations of the terms, and implies differential classroom practices affecting student OTL. Perhaps the strongest evidence for teachers' varying understandings about assessment practices coherent to standards lies in the item addressing *alignment of classroom tests* with the mathematics content standards. Although teacher self-ratings correlated with authentic classroom assessment practices reported, this was significant only at the secondary level, and practical significance was negligible; teacher ratings of their classroom assessment practices only predicted a little over 16% of the variance in authentic assessment practices.

A factor that likely contributes to this mismatch between perceived and actual alignment of assessment practices with reform recommendations is the generally low level of investment in reform-targeted professional development. Although teachers report that they have participated, on average, in from 5 to 10 days' worth of professional development opportunities over the past three years that have been focused in general on mathematics standards, instruction, curriculum, and assessment, and specifically on assessments or performance assessments related to standards, these numbers likely do not reflect an adequate amount of time for teachers to develop full understanding and effective implementation strategies. Various assessment researchers (Aschbacher, 1993; Shepard, 1995) have emphasized the extensive amount of time needed for teachers to learn about and grow comfortable with new assessments, developing or reviewing and selecting them, using them in the classroom, to be trained in rating student work, to do scoring, and to synthesize assessment results to make instructional and program decisions.

Perhaps pivotal in this scenario has been the role that reform policy has taken in Colorado, *by omitting any specific conceptual links between the standards legislation and implied changes in instruction or assessment*. In order to avoid political battles erupting in the early 1990's around "outcomes-based" education, the 1993 legislation that introduced "standards-based education" into state law specifically defined "standards" in terms of "content standards." Consistent with the state's Constitutionally-protected tradition of local control, state policy makers have scrupulously avoided explicitly connecting standards to constructivist ideas about instruction or assessment. Standards are equivalent to *content*. Part of the related message that has gone out to many Colorado teachers is that they do *not* need to change instructional and assessment practices; rather, they simply need to make sure that they cover the newly defined *content* to be in compliance with the standards. In this sense, state policy itself has constrained the potential for implementing standards-based reform consistent with its full intent, although it is likely that local sites that already possessed high levels of capacity for this reform are going beyond the minimum implied by the state.

Consonant with state-level *laissez-faire* regarding the interpretation of standards reforms is a lack of attention to local capacity building. While districts were required to adopt "standards-based education" (at least in terms of educational terminology to be used around content objectives), no additional funding or considerations were made for related needs around professional development, local assessments, or different instructional materials. In Colorado, the state provides no time, neither does it provide any funding for professional development; thus the aspect of local capacity addressed by teacher training efforts is free to vary depending on local resources, without any state intervention or equalization. Additionally, local guidelines for professional development vary widely, and are frequently characterized by a "smorgasbord" approach, without a coherent focus. In terms of *resource capacity*, the state has also provided for little in the way of adequacy; education spending per student has decreased 4% (adjusted for inflation) from 1986 to 1996, although relative income has increased (Education Week, 1998). According to Smith, et al (1995), similarly inadequate capacity and inadequate approaches to capacity building impeded coherent responses to the Arizona assessment reform.

It should be noted that standards-related reforms rely largely on the notion of large-scale assessment as a lever for instructional change; hence the argument from reformers to improve the quality of assessments to drive instruction toward higher-order goals. Joan Herman reminds us that, "in the absence of serious teacher capacity building to support instructional improvement, pressure to improve test scores may well corrupt both the teaching and learning process and the meaning of the test scores." (Herman, 1997, p. 6)

These findings have serious implications for student OTL. Teachers not only report a variety of different practices in mathematics classrooms across the state, inconsistencies in their responses likely indicate considerable variation in teacher capacity for in-depth understanding and effective implementation of more progressive assessments. Additionally, given the lack of state level attention in Colorado to issues of capacity-building, these data indicate that the probable inequities in student OTL examined here will likely increase in magnitude. Students in classrooms with capacity-rich teachers, who are likely situated in capacity-rich schools and districts, may receive the opportunities to learn the thinking curriculum that they need to do well on upcoming assessments. However, those with less-well-prepared teachers, in poorer sites with fewer human capital resources, will likely suffer.

The results of this study have broad implications for further research. Information such as this from carefully designed and administered surveys can provide valuable insights about the current status of assessment practices in classrooms across the state and serve as one source of data about variations in student opportunities to learn the content of more complex assessments. However, it is necessary to conduct validation studies and to supplement and triangulate survey data with alternate sources of information (e.g., classroom observation, document review, analysis of assessment results, and interviews). It should be noted that, at best, this measure provides data about the relative proportion of reported teacher activities around classroom assessment practices. It does not provide any information about the quality of implementation, which can best be addressed through more direct measures.

Additionally, those resources and experiences that build greater local capacity for implementing and more fully realizing standards and assessment reforms need to be studied in greater depth, including further examination into the nature of teacher professional learning opportunities and how these relate to teacher practice and student achievement. Further study is also needed to better understand and operationalize student OTL, as part of a clarification of delivery standards.

References

American Association for the Advancement of Science. (1993). *Benchmarks for Science Literacy.* New York: Oxford University Press.

Aschbacher, P.R. (1993). *Issues in innovative assessment for classroom practice: Barriers and facilitators.* [CSE Technical Report 359]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Baron, J. B. (1990). Performance assessment: Blurring the edges among assessment, curriculum, and instruction. In A. B. Champagne, B.E. Lovitts, & B. J. Callinger (Eds.) *Assessment in the service of instruction* (127-148). Washington, DC: American Association for the Advancement of Science.

Blank, R.F. (1993, Spring). Developing a system of education indicators: Selecting, implementing, and reporting indicators. *Educational Evaluation and Policy Analysis, (15)*, 1. (65-80).

Burstein, L., McDonnell, L., Van Winkle, J., Ormseth, T., Mirocha, J., & Guiton, G. (1995). *Validating national curriculum indicators.* Santa Monica, CA: RAND.

Center for Research on the Context of Teaching (1994). *Survey of elementary mathematics education in California.* Stanford, CA: Stanford University.

Clune, W. (1995). Educational adequacy: A theory and its remedies. *University of Michigan Journal of School Law, 28*(3), 481-491.

Cohen, D.K. (1990). Revolution in one classroom. In S. H. Fuhrman & B. Malen (Eds.) *The politics of curriculum and testing.* (103-123). New York: The Falmer Press.

Colorado Department of Education (1995). *Colorado State Model Content Standards in Mathematics.* (Adopted 6/8/96). Available: http://www.cde.gov. (Accessed 4/2/98).

Conference Report on H.R. 1804, Goals 2000: Educate America Act.. (1994). Washington, D.C.: Congressional Record, 140(32), H1625-H1684.

Conley, S., & Odden, A. (1995, Summer). Linking teacher compensation to teacher career development. *Educational Evaluation and Policy Analysis(17)*, 2. (219-237).

Darling-Hammond, L. (1995). Equity issues in performance based assessment. In M. T. Nettles & A. L. Nettles (Eds.) *Equity and excellence in educational testing and assessment* (89-114). Boston, MA: Kluwer.

Education Week. (1998, January 8). The state of the states. In *Quality Counts*. Special Edition.

Educational Testing Service (1992). *National assessment of educational progress: Grade eight mathematics teacher questionnaire*. Plainfield, NJ: Author.

Elmore, R.F. (1996, Spring). Getting to scale with good educational practice. *Harvard Educational Review(66)*, 1. (1-26).

Firestone, W.A. (1989). Educational policy as an ecology of games. *Educational Researcher (18)*, 7. (18-24).

Frederiksen, J. R., & Collins, A. (1989, December). A systems approach to educational testing. *Educational Researcher (18)*, 9.(27-32).

Haladyna, T. M., Nolen, S. B., & Haas, N.S. (1991). Raising standardized achievement test scores and the origin of test score pollution. *Educational Researcher (20)*, 5. (2-7).

Herman, J. L., Klein, D. C. D., & Wakai, S. T. (1996). *Assessing equity in alternative assessment: An illustration of opportunity-to-learn issues*. [CSE Technical Report 440] Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Herman, J. L. (1997). *Large-scale assessment in support of school reform: Lessons in the search for alternative measures*. [CSE Technical Report 446]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Hiebert, E. H., & Raphael, T. E. (1996). Assessment in early literacy classrooms. In *Early literacy instruction*. (Pp. 173-212). New York: Harcourt Brace Jovanovich.

International Association for Evaluation of Educational Achievement. (1994). *Third international mathematics and science study: Teacher questionnaires, populations 1, 2, and 3*. The Hague, The Netherlands: Author.

Jennings, N.E., & Spillane, J.P. (1996). State reform and local capacity: Encouraging ambitious instruction for all and local decision making. *Journal of Educational Policy(11)*, 4. (465-482).

Lomax, R. G., West, N. M., Harmon, M.C., Viator, K. A., & Madaus, G. F. (1992). *The impact of mandated standardized testing on minority students*. Boston: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Madaus, G. F. (1991). The effects of important tests on students: Implications for a national examination system. *Phi Delta Kappan(73)*, 226-231.

McDonnell, L. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis, 17*(3), 305-322.

McLaughlin, M. W., & Shepard, L. A. (1995). *Improving education through standards-based reform*. Stanford, CA: The National Academy of Education.

Muthen, B., Huang, L., Jo, B., Khoo, S., Goff, G., Novak, J., & Shih, J. (1995). Opportunity-to-learn effects on achievement: Analytical aspects. *Educational Evaluation and Policy Analysis, 17*(3), 371-403.

National Center for Education Statistics (1993). Schools and staffing survey: Public school teacher questionnaire. In *1990-91 schools and staffing survey: Data file user's manual* (NCES Publication No. 93-144-1, pp. N1-N22). Washington, DC: Author.

National Center for Education Statistics (1996). *Pursuing excellence: Initial findings from the Third International Mathematics and Science Study (TIMSS).* US Department of Education. Available WorldWideWeb: http://www.ed.gov/NCES/timss/97198.html

National Council on Education Standards and Testing. (1992). *Raising standards for American education.* Washington, D.C.: U.S. Government Printing Office.

Prestine, N.A., & McGreal, T. L. (1997, August). Fragile changes, sturdy lives: Implementing authentic assessment in schools. *Educational Administration Quarterly (33)*, 3 (371-400).

Resnick, L. B., & Resnick, D.P., (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M.C. O'Connor (Eds.) *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer.

Rottenberg, C., & Smith, M.L. (1990, April). *Unintended effects of external testing in elementary schools.* Paper presented at the annual meeting of the American Educational Research Association, Boston.

Shanker, A. (1994). National standards. In C. Finn & H. Walberg (Eds.), *Radical education reforms* (pp. 3-20). Berkeley, CA: McCutchan Publishing Corporation.

Shepard, L.A. (1989). Why we need better assessments. *Educational Leadership.*

Shepard, L.A. (1995, February). Using assessment to improve learning. *Educational Leadership (52)*, 5. (38-53).

Smith, M. L. (1994, September). *How assessments work: Lessons learned in equity.* Presentation at the annual conference of the National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles.

Smith, M. L., Noble, A.J., Heinecke, W., Seck, M., Parish, C., Cabay, M., Junker, S.C., Haag, S., Tayler, K., Safran, Y., Penley, Y., & Bradshaw, A. (1997). *Reforming schools by reforming assessment: Consequences of the Arizona Student Assessment Program (ASAO): Equity and teacher capacity building.* [CSE Technical Report 425]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Smith, M., & O'Day, J. (1990). Systemic school reform. In S. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-267). New York: The Falmer Press.

Spillane, J. P. (1994). How districts mediate between state policy and teachers' practice. In R. F. Elmore & S.H. Fuhrman (Eds.), *The governance of curriculum.* Alexandria, VA: Association for Supervision and Curriculum Development.

Spillane, J. P. (1998, February). A cognitive perspective on the role of the local educational agency in implementing instructional policy: Accounting for local variability. *Educational Administration Quarterly(34)*, 1 (31-57).

Spillane, J. P., & Thompson, C. (1997, Summer). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis(19)*, 2 (185-203).

Takahira, S., Gonzales, P., Frase, M., & Salganik, L. H. (1998). *Pursuing excellence: A study of U. S. Twelfth-grade mathematics and science achievement in international context.* Washington, DC: National Center for Education Statistics.

TIMSS International Study Center (1996, November). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

TIMSS International Study Center (1997, June). *Mathematics achievement in the elementary school years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Voigt, J.(1995). Thematic patterns of interaction and sociomathematical norms. In Cobb, P., & Bauersfeld, H. (Eds.) *The emergence of mathematical meaning: Interaction of classroom cultures.* Hillsdale, NJ: Erlbaum Associates.

Weatherley, R., & Lipsky, M. (1977). Street-level bureaucrats and institutional innovation: Implementing special-education reform. *Harvard Educational Review (47)*, 2.(171-197).

Weiss, I.R. (1993). *National survey of science and mathematics education.* Chapel Hill, NC: Horizon Research, Inc.

Weiss, I.R. (1994). *A profile of mathematics and science education in the United States.* Chapel Hill, NC: Horizon Research.

Wiggins, G. (1989, April). Teaching to the (authentic) test. *Educational Leadership.*

Winfield, L. F., & Woodard, L. D. (1994). *Assessment, equity, and diversity in reforming America's schools.* [CSE Technical Report 372]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Wolf, D. P. (1993). Assessment as an episode in learning. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (213-240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wolf, D.P., & Reardon, S.F. (1996). Access to excellence through new forms of student assessment. In J.B. Baron & D.P. Wolf (Eds.) *Performance-based student assessment: Challenges and possibilities.*(1-31). Chicago, IL: University of Chicago Press.

TM029179

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Mathematics Assessment Practices in Colorado Classrooms: Implications About Variations in Capacity and Students' Opportunities to Learn	
Author(s): Ravay L. Snow-Renner	
Corporate Source: Colorado Educational Policy Consortium, University of Colorado at Denver	Publication Date: April, 1998

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Ravay Snow-Renner	Printed Name/Position/Title: Ravay Snow-Renner	
Organization/Address: CEPC/UCD 1444 Wazee St, Ste 230, Denver Co 80202	Telephone: 303-620-4090	FAX: 303-620-4588
	E-Mail Address: rsnow@cu.campus.mci.net	Date: 8/17/98

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on Assessment and Evaluation
University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com