Assessing Opportunity to Learn:
A California Example

CSE Technical Report 453

Joan L. Herman and Davina C. D. Klein

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)

December 1997

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein was supported in part under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA Catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

ASSESSING OPPORTUNITY TO LEARN: A CALIFORNIA EXAMPLE

Joan L. Herman and Davina C. D. Klein
National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

Across the nation, states, districts, and schools are working to transform American education by setting rigorous academic standards for student performance and establishing assessment systems to help assure that all students achieve these standards. Targeted on the complex thinking and problem-solving skills students will need for future success, the policy goals are ambitious and reflect a dramatic change in the expectations for student performance and in the curriculum and instruction in which students are to be engaged. In many cases, the stakes are equally dramatic: Those who do not meet expectations—those who do not pass the assessments—may be held back and/or not allowed to graduate; those who do pass will be granted important access. The potential consequences for these students' futures are substantial indeed. Are these policies working? Are these policies fair for all students? What other actions are necessary to support the objectives of such policies? The answers to these questions require accurate information about the quality of curriculum and instruction available to students—the opportunities that schools provide students to learn what is expected of them. In the current lexicon, this is termed students' *opportunity to learn* (OTL). In this paper, we explore the "what" and "why" of assessing opportunity to learn and then illustrate some of the challenges of assuring accurate measurement, using data collected as part of a pilot study of eighth-grade mathematics for a statewide assessment system.

Background

What is "Opportunity to Learn"?

Building from prior work on education and quality indicators (Blank, 1993), the OTL concept operationalizes what is taking place in schools and classrooms to support students' learning and progress, particularly relative to new expectations for student performance being formulated at the national, state, and local levels. It asks: Are the curriculum and pedagogy in which students are engaged appropriate to the rigorous standards that have been established for student performance? In short, are students being provided opportunities to learn that which is expected of them—especially that for which they are held accountable in new assessment systems?

In examining such opportunities, OTL researchers typically have distinguished three overlapping categories of concern: curriculum content, instructional strategies, and instructional resources (Brewer & Stacz, 1996). The issue in curriculum content is the extent to which students have been exposed to the specific subjects and topics that are essential to attaining particular standards and/or that are directly assessed. We include not only domain-specific information (such as a mathematical formula to calculate the perimeter of a rectangle, or the meaning of fractional values), but also a deeper understanding of the ways of thinking in the domain being studied. How well do students understand the connections between pieces of factual information? How much do students understand the "hows" and "whys" rather than just the "whats" and the procedural knowledge involved in mathematics? How often do students work on mathematical problems that require complex thinking or problem-solving skills, if such skills are the target of standards and assessments?

In looking at instructional strategies, we ask whether students have been exposed to the kinds of teaching and instructional experiences that would prepare them for success. At the most elemental level, do students have experience with the kinds of tasks by which their performance is to be judged? For example, for alternative assessments that employ nonstandard questions, are students familiar with open-ended items? To what extent have they been asked by their classroom teachers to explain their reasoning and justify their responses? Have they worked with varied representational systems, such as graphs, charts, and tables, and used these systems to complete an assessment? More generally, have students been engaged in the kinds of constructivist educational processes that current theory holds essential for maximum student development and for achievement of high levels of complex performance (e.g., active student engagement, opportunities to create new meaning and to link existing knowledge to new knowledge, active exploration and collaboration, problems that elicit critical thinking skills)?

Instructional resources form the third category of OTL interest: Are there appropriate resources to prepare students for success on standards? Prime within this category are issues of teacher preparation, such as education (e.g., college degree), amount and type of teaching experience, participation in relevant inservice education, and attitudes towards mathematics instruction. We expect that in order to implement the reform agenda in mathematics—to pose appropriate, real-world mathematical problems; to elicit deeper mathematical thinking; to portray mathematics as relevant, interesting, and useful in today's world—teachers must have a significant grounding in the discipline of mathematics. They must be able to think mathematically and scientifically; they must be able to reason about the concepts, principles, and procedures used in this discipline. Although professional development may help those teachers who already have a strong mathematical background, teachers need to have a working knowledge and a sophisticated understanding of mathematics in order to teach to rigorous standards.

Material resources also are relevant to this category, such as appropriate instructional materials (recent textbooks and supplementary materials) and suitable tools (e.g., calculators, manipulatives, other technology). Just as mathematical background is important, support materials in the form of recent textbooks and classroom tools are essential in the innovative teaching of mathematics. We cannot expect teachers without a great deal of mathematics background, without up-to-date support materials, and without the appropriate classroom tools to be able to transform the teaching of mathematics into a hands-on, relevant, and interesting experience for students. All of these resources are an important part of evaluating students' opportunity to learn.

The Importance of Assessing Opportunity to Learn

Along with other researchers, we argue that OTL measures provide policy makers with critical information on how their policies are operating and essential feedback on whether assumptions underlying the policy are accurate (e.g., see McDonnell, 1995; Porter, 1995). The assumptions and chain of logic underlying these policies appear relatively simple and straightforward: Assessment can effectively communicate educational goals and expectations for student performance. With suitable incentives and sanctions (either external or internal), teachers and schools will respond to a new assessment by changing their curriculum and instruction to provide students with appropriate opportunities to learn what is expected. Students too, encouraged by the threat of adverse consequences or the promise of future reward, will take the standards seriously and will be motivated to learn what is expected of them. All in the system will take feedback about performance seriously and use that information to improve what they do.

Implicit in this presumed cycle of assumptions are at least three points where OTL data can serve the critical functions of verifying the policy assumptions and contributing to system decision making. First, central to the system logic is the notion that school curriculum, teaching, and instruction will change to provide students with appropriate opportunities to learn and to achieve the high standards that are held for them. Absent such changes, improvement in student performance is unlikely and the policy tool ineffective in achieving its ends. Yet, absent data on OTL, how will policy makers know?

Second, because there often are important consequences attached to test performance that can dramatically impact students' futures, fairness demands that all students be provided with appropriate opportunities to achieve desired standards. Policies may provide students the motivation to achieve, but unless the educational system does its job in providing educational opportunities, students will be unable to perform at expected levels. Research shows great inequity in distribution of educational resources and access to knowledge: Economically disadvantaged and culturally diverse subgroups have had less access than other students to a challenging curriculum that would prepare them for success on today's standards (Guiton & Oakes, 1995). Students cannot and should not be held accountable for knowledge and skills they have had no opportunity to acquire. Absent data on OTL, how can policy makers and the public verify that all students have a chance to succeed, and that all subgroups of the population have the opportunity to engage in the kinds of curriculum and instruction that would prepare them to achieve expected standards?

In addition to serving these two essential policy purposes, data on students' OTL can provide important feedback to schools to stimulate their thinking about the strengths and weaknesses of their curriculum and course offerings and about their priorities for professional development, materials acquisition, and resource allocations. High-stakes uses of OTL data have been proposed to hold schools and teachers accountable for their educational practices and for adherence to particular standards of educational delivery (McDonnell, 1995; National Council on Education Standards and Testing, 1992).

Finally, and as alluded to above, OTL data can provide policy makers with early feedback on system progress. Studies of school reform document well the time required to change practice and the challenge of improving student performance. Immediate, dramatic improvements in student performance are not a realistic expectation, particularly when sophisticated changes in curriculum and instruction must precede such improvement. In the meantime, how might policy makers know whether the system is going in the right direction? We would argue that data on students' opportunity to learn can provide an interim measure of system progress as well as important data to inform midcourse corrections.

There are innumerable other potential and important uses of OTL data—in studies of the instructional sensitivity of tests, to condition test results, or in research on curriculum, pedagogy, and effective strategies, for example—but the points raised above should be sufficient to establish the importance of reliable OTL measures for policy purposes.

Difficulties in Assessing Opportunity to Learn

Given the importance of assessing OTL, policy makers and educators have nonetheless had difficulty developing valid and useful OTL indicators (Guiton & Oakes, 1995). Although almost everyone seems to agree that OTL indicators should describe the resources, school conditions, curriculum, and instruction to which students have access (Burstein, 1993; Guiton & Oakes, 1995; Wiley & Yoon, 1995), there clearly are many ways to structure these measures. Methods for routine collection of such data (e.g., in regular program monitoring or as regular parts of state or local assessment systems), as well as the costs and efficiency of data collection, are clearly at issue.

One of the difficulties in assessing students' opportunity to learn is discovering how (and from whom) it is best to collect OTL data. The Third International Mathematics and Science Study has taken the most comprehensive approach to data collection to date, including detailed teacher and student surveys, observations, and coding of instructional materials (Schmidt & McKnight, 1995). A number of other policy studies examining quality indicators and school reform have combined a variety of data collection strategies, including survey, materials analysis, teacher logs, transcript review, and direct observation (Burstein, McDonnell, Van Winkle, Ormseth, Mirocha, & Guiton, 1995; Jones, Davenport, Bryson, Bekhuis, & Zwick, 1986; McDonnell, 1995; Porter, 1991). But, no doubt for reasons of cost, most local, state, and national assessments of OTL have utilized only surveys of practice, based on teacher and/or student reports (Brewer & Stacz, 1996; Smithson, Porter, & Blank, 1995). Do such reports provide accurate information about classroom opportunities? Are the responses reliable? Does it matter whether teachers or students are queried? As part of a larger planned study of the implementation and impact of the California state assessment system, we collected multiple measures of students' OTL that enabled us to investigate such questions.

Methods

Our data were collected in conjunction with the 1993 California Learning Assessment System (CLAS) Middle Grades Mathematics Performance Assessment (California Department of Education, 1993). In this section, we describe participants, materials, and data collection efforts as they relate to this exploration; refer to Herman, Klein, Heath, and Wakai (1995) for more in-depth coverage of our methodology.

Participants

Thirty-six classrooms, 27 teachers, and over 800 eighth-grade students from 13 schools across California participated in our study. Because our primary interests centered on equity, inner-city schools reflecting economically disadvantaged and culturally diverse communities were overrepresented. Within each school, three classrooms were selected to represent the range of math classes available at the school (e.g., algebra, pre-algebra, eighth-grade mathematics). Schools in our sample were part of a larger volunteer sample involved in a pilot study conducted by the California Department of Education.

Materials

The 1993 CLAS Mathematics Performance Assessment, although now defunct, was an example of an alternative assessment designed to measure students' complex mathematical thinking, communication, and problem-solving skills. The assessment featured a matrix-sampled design with eight different assessment forms, each containing two open-ended questions and eight multiple-choice items. In addition, all students in our study (as part of a special pilot study by the state) completed a "common form" of the assessment on the day following the regular statewide administration. Thus, our data collection focused on students who had all completed the *same* assessment (as opposed to one of eight forms). As in all of the forms, the two open-ended tasks were designed to pose authentic, relevant problem situations for students to solve, asking students to explain their responses and to construct those responses using multiple modes of representation. The multiple-choice items were intended to assess mathematical thinking.

Data Collection

Our data collection efforts utilized both student surveys and teacher interviews to gather information on areas such as classroom instructional practices, direct preparation for the CLAS, content preparation, teacher background and experience, and resource quality. Some instructional artifacts (assignments and tests) also were gathered from teachers during the interviews and were later coded for evidence of reform practices. In addition, we planned to use teacher data collected by the state upon which we intentionally modeled many of our student survey items. However, we were unable to link these state-collected teacher data to our data set because teachers and their classes were not individually identified in the state data. Data on student performance similarly were not available. Thus, an unfortunate sequence of events limits both the number of comparisons we are able to make between teacher and student reports as well as our ability to link these reports to classroom performance. Though imperfect as a representative sample and incomplete in its available data, our study nonetheless serves as a point of departure for examining important validity issues in assessing students' opportunity to learn.

Results

Opportunity-to-Learn Variables

Based on preliminary data analysis, seven OTL variables were constructed from our teacher interview data and six were constructed from the student survey data, including both composite and single-item variables. Variables were selected to represent various aspects of the OTL construct—curricular content, instructional strategies, direct preparation for the CLAS, and quality resources—including those aspects that might be expected to be highly related and those for which there was little expectation of convergence. In addition, analysis of the classroom artifact data, including classroom assignments and tests, provided an independent measure of classroom instructional opportunities. (See Table 1 for a list of OTL variables used.) All data were aggregated to the classroom level because this is the level at which comparisons between student and teacher reports are possible.

Table 1

Opportunity-to-Learn Measures

Name of OTL variable (Category)	Definition of OTL variable	Source of variable
REFRMTCH (specific practices)	Reform classroom instructional practices 7 subvariables included, $\alpha = .64$	Teacher interview
GENALIGN (general practices)	General instructional alignment with CLAS 2 subvariables included, $\alpha = .68$	Teacher interview
CONTENTT (content preparation)	Content preparation for specific math topics included on CLAS	Teacher interview
PREPTIME (direct preparation)	Class periods spent using CLAS Sampler	Teacher interview
MATHBACK (resources: teacher)	Mathematics background of teacher 4 subvariables included, $\alpha = .51$	Teacher interview
NCALC (resources: tools)	Number of calculators available in the classroom	Teacher interview
RECENT (resources: textbook)	Recency of mathematics textbook used	Teacher interview
RFRMSTUD (specific practices)	Reform classroom instructional practices 11 subvariables included, $\alpha = .77$	Student survey
GENPREP (general preparation)	General impression of preparation for CLAS	Student survey
CONTENTS (content preparation)	Content preparation for specific math topics included on CLAS	Student survey
CALCHOME (resources: tools)	Availability of calculators at home	Student survey
RFRMTMAT (specific practices)	Reform classroom instructional practices 11 subvariables included, $\alpha = .94$	Teacher materials

Curriculum content. Information about curricular topic coverage was solicited from teachers and students. Teachers were asked how much class time they had spent on each of six specific mathematical topics (e.g., area, fractions) covered by the CLAS common form (composite scale: CONTENTT). A similar question about the same six topics was asked of the students ("How well have your math classes prepared you to answer questions in the following areas?"; composite scale: CONTENTS).

Instructional strategies. Both teacher and student variables assessed the extent to which students were engaged in the types of activities that were likely to prepare them to do well on assessments of complex mathematical thinking and problem solving, based on mathematics reform literature (National Council of Teachers of Mathematics, 1991). The teacher-based "reform instructional practices" composite scale (REFRMTCH) was created from seven items that asked teachers how often they engaged in particular classroom practices (e.g., written math assignments, extended projects, small group work; see Table 2 for a complete list). Responses to these items showed reasonable internal consistency with a Cronbach's alpha coefficient equal to 0.64. Student survey items asked students similar questions regarding the frequency of various types of classroom activities. Eleven items of the form "In your current math class, how often do you . . ." were combined to form a student-reported reform practices scale (RFRMSTUD). These too showed reasonable internal consistency (α = .77; see Table 3 for a listing of these items).

Table 2

Items Included in the Teacher-Reported Reform Instructional Practices Scale

Name of OTL variable	Definition of OTL variable
likeclas	How often are students asked to solve open-ended items similar to those on the CLAS (i.e., no obvious solutions, no single correct strategy, and students must explain their reasoning)?
Written	How often do you ask your students to reflect upon and explain their mathematical reasoning in written assignments?
smgroup	How often do you ask your students to reflect upon and explain their mathematical reasoning in small groups?
clasdisc	How often do you ask your students to reflect upon and explain their mathematical reasoning as part of class discussions?
mathport	Do your students keep math portfolios?
intass	Rating of "integratedness" of major assignments
extass	Rating of "extendedness" of major assignments

Table 3

Items Included in the Student-Reported Reform Instructional Practices Scale

Name of OTL variable	Definition of OTL variable ("In your current math class, how often do you . . .")
WORDPROB	Practice doing word problems?
MORE1WAY	Practice doing problems that can be solved in more than one way?
REALTHNK	Practice doing problems you have to really think about to come up with a solution?
EXPNTHNK	Practice doing problems that ask you to explain (in writing) your thinking?
MATHPROJ	Do math projects that take at least a week to complete?
REALLIFE	Apply math to real-life practical problems?
SMALLGPS	Work in small groups to answer problems?
CALCPROB	Use calculators to solve problems?
MANIPUL	Use rulers, blocks, or solids to explore problems?
ORALPRES	Give an oral presentation?
WRTEPARA	Answer math problems that require you to write a paragraph or more?

Classroom materials from each teacher also were coded for evidence of the types of instructional engagement queried on the survey. The composite scale (RFRMTMAT) so created included 11 items ($\alpha = .94$; see Table 4).

Table 4

Items Included in the Teacher Materials Reform Instructional Practices Scale

Name of OTL variable	Definition of OTL variable
APPLYNEW	Items requiring application of concept to new situation
NOOBVMTD	Items with no obvious method of solution
MORE1ANS	Items with more than one possible answer
MORE1APP	Items with more than one possible approach
XPLNRSNG	Items requiring explanation of reasoning
WRITE1P	Items requiring at least one paragraph of writing
REALAPPS	Items involving real-life applications
CRTQSOLN	Items requiring critique or analysis of given solution
DESCHOW	Items requiring description of how to solve problem
EXTENDP	Items requiring gathering data/conducting experiments/extended investigations
OVERALLR	Overall general math reform rating

General alignment of instructional practices. A more general measure of alignment of classroom practices (GENALIGN) was created using two highly correlated teacher-reported items ($\alpha = .68$). These two items asked teachers to rate (a) how aligned their classroom instruction was with the new CLAS-type assessments, and (b) how well prepared they were to teach to CLAS-type objectives. From the student perspective, a general impression of CLAS preparation measure (GENPREP) was obtained by asking students how much they agreed with the statement "I was well prepared to take the CLAS." In contrast to the specific items described above, these OTL measures seek to capture classroom culture with more general questions.

Direct preparation for the CLAS. We asked teachers how much class time they had spent working directly with the CLAS Sampler or Addendum (California Department of Education, 1992), a set of sample CLAS-type exercises disseminated by the California Department of Education (variable: PREPTIME).

Quality instructional resources. Teachers' mathematical background served as one measure of resource quality. Our measure of mathematics background (MATHBACK) was an aggregate based on four items from the teacher interview: college major/minor, type of teaching credential held, amount of college coursework, and inservice training in mathematics. The composite was weighted to recognize the importance of major college preparation in math or a math-related field.

In addition to teachers as resources, we looked at two other measures of classroom resources: textbook recency as a proxy for quality, and calculator availability. Clearly, in an environment in which new assessments demand new ways of teaching, old textbooks which do not feature reform goals may well hinder the learning process. Teachers were asked to indicate the primary textbook used in their classroom; these were then coded for recency and thus likelihood of fit with reform curriculum (RECENT). Finally, calculators were taken be to a proxy for the availability of mathematical tools and manipulatives—an important part of an innovative curriculum. Both teacher-reported data regarding the number of students with calculators (whether home- or school-provided; NCALC) and student-reported data regarding calculator accessibility at home (CALCHOME) were used.

Patterns of Relationships Among Opportunity-to-Learn Measures

Correlations among teacher-reported variables. Table 5 shows correlations among the various teacher OTL measures and the measure derived from coding classroom instructional artifacts. The patterns show some support for the consistency and general validity of teacher reports: Variables that would be expected to be positively related and those which would be expected to be unrelated in general are so. There is a significant correlation between the teacher-reported specific instructional practices scale and teacher reports of the general alignment of their classroom instruction to the CLAS (REFRMTCH and GENALIGN, $r = .37$, $p < .05$). Also, as would be expected because the CLAS practice items engage students in some of the specific instructional processes queried, direct preparation for the CLAS (i.e., time spent on practice exercises) is significantly correlated with the specific instructional practices scales (PREPTIME and REFRMTCH, $r = .51$, $p < .01$), although not with the general alignment scale (PREPTIME and GENALIGN, $r = .26$, $p > .05$). Finally, the content preparation scale is significantly and positively related to general alignment (CONTENTT and GENALIGN, $r = .54$, $p < .01$), but not to the teacher-reported specific instructional practices scale (CONTENTT and REFRMTCH, $r = .33$, $p > .05$), nor to direct preparation time (CONTENTT and PREPTIME, $r = .16$, $p > .05$).

Table 5

Correlations Among Teacher Opportunity-to-Learn Measures

	REFRMTCH	GENALIGN	CONTENTT	PREPTIME	RFRMTMAT	MATHBACK
REFRMTCH	–	.37*	ns	.51**	.76**	ns
GENALIGN	–	–	.54**	ns	.46**	ns
CONTENTT	–	–	–	ns	.48**	ns
PREPTIME	–	–	–	–	.59**	ns
RFRMTMAT	–	–	–	–	–	ns
MATHBACK	–	–	–	–	–	–

* $p < .05$, ** $p < .01$.

As would also be expected, given that teachers do not have control over district or school budgets for resources, classroom resources were not found to be significantly correlated with each other (MATHBACK and RECENT, $r = –.23$, $p > .05$; MATHBACK and NCALC, $r = –.16$, $p > .05$; RECENT and NCALC, $r = –.02$, $p >$

.05). Teacher background also is unrelated to any of the other teacher OTL measures, suggesting that teachers' content knowledge is independent of their philosophies of instruction and their pedagogical decision making.

Especially noteworthy is the high correlation between teacher self-reports and independent analysis of classroom instructional materials (REFRMTCH and RFRMTMAT, $r = .76$, $p < .01$; GENALIGN and RFRMTMAT, $r = .46$, $p < .01$), which could be regarded as a concurrent criterion measure. The materials scale also is positively related to teacher reports of content preparation (CONTENTT and RFRMTMAT, $r = .48$, $p < .01$). That time directly preparing students for the CLAS is significantly related to teachers' use of innovative curriculum materials (PREPTIME and RFRMTMAT, $r = .59$, $p < .01$) may suggest that teachers are incorporating the models provided by the CLAS practice activities into their routine classroom practices.

Correlations among student-reported variables. Turning next to student data (Table 6), we found a significant correlation between the student-reported instructional practices scale and the student general measure of preparation (RFRMSTUD and GENPREP, $r = .42$, $p < .05$); students who engaged in instructional practices more consistent with reform practices were more likely to report themselves well prepared to take the CLAS. The general preparation measure also was found to be significantly correlated with specific content preparation reports (CONTENTS and GENPREP, $r = .54$, $p < .01$). In addition, in contrast to teacher results, student reports of specific content preparation were significantly and positively related to their reports of innovative instructional strategies (CONTENTS and RFRMSTUD, $r = .34$, $p < .05$). Finally, as might be expected, the calculator availability measure (CALCHOME) was uncorrelated with the other student measures.

Table 6

Correlations Among Student Opportunity-to-Learn Measures

	RFRMSTUD	GENPREP	CONTENTS	CALCHOME
RFRMSTUD	–	.42*	.34*	ns
GENPREP	–	–	.54**	ns
CONTENTS	–	–	–	ns
CALCHOME	–	–	–	–

* $p < .05$, ** $p < .01$.

Teacher-student correlations. Having examined relationships within teacher and student measures, we now examine relationships between them. We focus first on those measures that were asked in similar form of both teachers and students: specific instructional strategies, general instructional alignment, and content coverage. As Table 7 shows, teacher-reported and student-reported items measuring specific classroom instructional practices were highly correlated (REFRMTCH and RFRMSTUD, $r = .63$, $p < .01$); student reports also were highly and positively correlated with the analysis of instructional artifacts (RFRMTMAT and RFRMSTUD, $r = .65$, $p < .01$).

Table 7

Correlations Among Student and Teacher Opportunity-to-Learn Measures

Teacher measures	Student measures			
	RFRMSTUD	GENPREP	CONTENTS	CALCHOME
REFRMTCH	.63**	ns	ns	ns
RFRMTMAT	.65**	ns	ns	ns
GENALIGN	.37*	ns	ns	ns
CONTENTT	.34*	ns	ns	ns
NCALC	ns	ns	ns	.53**
PREPTIME	.44*	.35*	ns	ns

* $p < .05$, ** $p < .01$.

In contrast, teacher and student responses to general alignment issues were not significantly correlated (GENALIGN and GENPREP, $r = .31$, $p > .05$). Teacher and student responses to our questions about content coverage also were completely uncorrelated (CONTENTT and CONTENTS, $r = .02$, $p > .05$). Even when looking at the correlations between individual items targeting each topic, teacher and student responses were not significantly related in five of the six areas; only with regard to distance/time problems do student and teacher reports show agreement ($r = .51$, $p < .01$).

Finally, even though assessing somewhat different contexts, calculator resource measures were significantly correlated across teacher and student reports (NCALC and CALCHOME, $r = .53$, $p < .01$).

Similar to patterns of relationships within teacher and student reports, results also show consistency in the two groups' responses about related aspects of OTL. The student-reported instructional practices scale is significantly correlated with the teacher-reported general alignment measure (RFRMSTUD and GENALIGN, $r = .37$, $p < .05$), with the teacher-reported content preparation scale (RFRMSTUD and CONTENTT, $r = .34$, $p < .05$), and with the teacher-reported direct preparation measure (RFRMSTUD and PREPTIME, $r = .44$, $p < .05$). In addition, the student-reported general preparation measure is significantly correlated with the teacher-reported direct preparation for the CLAS measure (GENPREP and PREPTIME, $r = .35$, $p < .05$).

Discussion and Conclusion

Patterns of relationships among and between teacher and student reports show areas of promise and challenge in using surveys to measure students' opportunity to learn. Teacher and student reports themselves show reasonable internal consistency, and patterns of relationships among various aspects of OTL are generally as would be expected: Related aspects show positive relationships, while aspects that logic suggests are unrelated are not related (e.g., the relationship between availability of recent textbooks and teachers' instructional strategies).

Further, it appears that students and teachers agree in their responses to some aspects of OTL: Measures of instructional practices and calculator use were highly correlated across respondent groups, and the student-reported instructional practices measure was found to be significantly correlated with a number of teacher-reported OTL measures. Importantly, both teacher and student reports of instructional strategies were strongly and positively related to the results of the independent analysis of instructional artifacts.

On the other hand, responses to content coverage items showed no consistency between the two respondent groups—an unexpected result. While the two sets of items were intended to solicit parallel information, hindsight indicates they functioned otherwise because of differences in the way in which the questions were phrased. Whereas teachers were asked about the amount of classroom instructional time spent on various topic areas, students were asked instead to rate how well their classes prepared them for questions in these topic areas. We were interested in obtaining an objective, classroom-level measure of topic preparation. Students, however, may have given us a more personal, subjective "how-well-prepared-I-feel" response, and patterns of relationships with socioeconomic status and with student performance reported elsewhere (Herman & Klein, 1996) indeed suggest that this is the case. Students' interpretation of the question appears different from the meaning we intended as survey developers, a clear illustration of the subtle ways in which how questions are asked can affect OTL results.

Another such example is reflected in our measures of general preparation and alignment. Students were asked how well prepared they felt to take the CLAS, and this was intended as a general measure of curricular alignment. However, like the topic-preparation items, this survey item may have tapped more into students' personal levels of academic efficacy rather than the nature of their classroom teaching and learning experiences. That is, students who feel confident in their mathematics ability—regardless of their specific content exposure, their assessment experience, or their classroom activities—may have responded more positively to this item than other students. Interestingly, however, both student-reported content preparation and general preparation measures did correlate significantly with the student-reported measure of specific instructional practices, as well as with each other.

These results may suggest that students have a rather undifferentiated view of their opportunities to learn. Alternatively or concurrently, it may mean that students are responding to what for them is a unitary concept: the quality of effective learning, regardless of specific opportunities to which they have been exposed. That is, whereas teachers were asked how much time they had spent exposing students to particular topics, students essentially were asked about the quality and effectiveness of that exposure. Perhaps implicit in student responses is the critical insight that being exposed to content is not the same as learning it. In fact, it is the quality and effectiveness of instructional opportunities, not the presence, absence, frequency, or duration of specific topics or particular instructional strategies, that are essential to student learning and improved student performance. Yet attention to such quality is missing from existing measures of OTL.

In summary, our results suggest that we can use relatively quick, efficient, and inexpensive surveys of either students or teachers to obtain accurate data on what is going on in classrooms and on the broad nature of the teaching and learning activities in which students have been engaged. This is not to say, of course, that more intensive data do not provide a richer, more comprehensive portrait and may be needed periodically to validate and supplement survey results (e.g., see Burstein et al., 1995). From survey results, however, we can differentiate the extent to which instructional reform is being implemented in classroom practice, at least at a surface level. This is decidedly good news.

The meaning and interpretation of results on content preparation are less clear since teacher and student reports were unrelated, and we had neither an independent measure of content exposure nor classroom performance data on specific topics to help us discern the relative accuracy of teacher and student reports. What is clear, however, is that what is in respondents' minds as they read and interpret survey items may not mirror the intentions of item writers. Expert review and simple trials with students, to which the surveys reported here were subjected, are not sufficient to assure that intended meanings remain intact. Instead, think-aloud protocols and try-outs of alternative wordings may be necessary to validate students' interpretations of OTL items and to assure accurate inferences from survey results. This is particularly crucial because concerns for student and teacher time limit the number of questions that can be asked and because, due to intended policy uses, results may have important consequences for teachers, schools, and the students within them.

Finally, that teachers and students provided accurate reports of classroom practice for our study, a decidedly low-stakes condition, does not necessarily mean they will do so under conditions with higher stakes. The history of testing suggests, in fact, that when accountability stakes are high, results can become corrupted (Herman, 1992; Koretz, 1988). The same policies that give rise to the current interest in assessing OTL contain within them the potential for misuse and corruption of OTL data. The policy uses of such data as well as validation research to assure the integrity and accuracy of measures merit continuing scrutiny.

References

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

Brewer, D. J., & Stacz, C. (1996). *Enhancing opportunity to learn measures in NCES data*. Santa Monica, CA: RAND.

Burstein, L. (1993). *Validating national curriculum indicators: A conceptual overview of the RAND/CRESST NSF Project*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Burstein, L., McDonnell, L. M., Van Winkle, J., Ormseth, T., Mirocha, J., & Guiton, G. (1995). *Validating national curriculum indicators* (DRU-1086-NSF). Santa Monica, CA: RAND.

California Department of Education. (1992). *Addendum to a Sampler of Mathematics Assessment*. Sacramento: Author.

California Department of Education. (1993). *The California Learning Assessment System*. Sacramento: Author.

Guiton, G., & Oakes, J. (1995). Opportunity to learn and conceptions of educational equality. *Educational Evaluation and Policy Analysis*, *17*, 323-336.

Herman, J. L. (1992). What research tells us about good assessment. *Educational Leadership, 49*(8), 74-78.

Herman, J. L., & Klein, D. C. D. (1996). Assessing equity in alternative assessment: An illustration of opportunity-to-learn issues. *Journal of Educational Research, 89*, 246-256.

Herman, J. L., Klein, D. C. D., Heath, T. M., & Wakai, S. T. (1995). *A first look: Are claims for alternative assessment holding up?* (CSE Tech. Rep. No. 391). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Jones, L. V., Davenport, E. C., Bryson, A., Bekhuis, T., & Zwick, E. (1986). Mathematics and science test scores as related to courses taken in high school and other factors. *Journal of Educational Measurement*, *23*, 197-208.

Koretz, D. (1988). Arriving in Lake Wobegon: Are standardized tests exaggerating achievement and distorting instruction? *American Educator, 12*(2), 8-15.

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

National Council on Education Standards and Testing. (1992). *Raising standards for American education: A report to Congress, the Secretary of Education, the National Education Goals Panel, and the American People.* Washington, D.C.: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston. VA: Author.

Porter, A. C. (1991). Creating a system of school process indicators. *Educational Evaluation and Policy Analysis*, *13*(1), 13-29.

Porter, A. C. (1995). The uses and misuses of opportunity to learn standards. *Educational Researcher, 24*(1), 21-27

Schmidt, W. H., & McKnight, C. C. (1995). Surveying educational opportunity in mathematics and science: An international perspective. *Educational Evaluation and Policy Analysis, 17*, 337-354.

Smithson, J. L., Porter, A. C., & Blank, R. K. (1995). *Describing the enacted curriculum: Development and dissemination of opportunity to learn indicators in science education*. Washington, D.C.: Council of the Chief State School Officers.

Wiley, D. E., & Yoon, B. (1995). Teacher reports on opportunity to learn: Analyses of the 1993 California Learning Assessment System (CLAS). *Educational Evaluation and Policy Analysis, 17*, 355-370.