Multiple Measures, Multiple Uses

Margaret E. Goertz
CPRE/University of Pennsylvania

2011 Reidy Interactive Lecture Series
Center for Assessment
September 22, 2011

The purpose of this presentation is to frame a discussion of multiple measures by looking at the (1) multiple **uses** of data in (2) different **levels** of the education system (student, teacher, school) and (3) the **types** of measures that can address each use. I use the term "measure" broadly to talk about types of data collected from different kinds of assessments (e.g., informal assessments, standardized assessments, grades, etc.) rather than specific ways of measuring student knowledge (e.g., test constructs, how grades are calculated, etc.).

This presentation has three parts: (1) an overview of potential uses of data by level of education system (Table 1); (2) a brief description of the types of measures that can be used to meet each purpose at the student, teacher and school levels, respectively (Tables 2-4); and (3) a discussion of the similarities and differences in types of measures for common purposes across the three levels and the issues this raises for the multiple use of measures (Table 5).

I make four major points in this presentation:

1. There has been an expansion of uses of measures over time (most recently teacher and principal evaluation and accountability).

2. Different uses (e.g., continuum from diagnosis to external accountability) embody different technical requirements for measures and how they are combined (e.g., standardization, reliability, validity).

3. Different uses call for data collection at different points in time (e.g., on-going monitoring of student work for student diagnosis versus more periodic assessment for informing teaching/learning versus summative measures for external accountability).

4. Aggregation of data across purposes within a level of the system (e.g., aggregation of student-level data as one moves from student diagnosis to student accountability) and across levels of the system (aggregation of student-level data at the classroom level to inform teaching and learning at the teacher and school levels) also raises a set of design and technical issues. What can be aggregated? What are the challenges of doing so?

Overview of potential uses of data by level of the education system (Table 1).

Table 1 shows different ways in which we currently use data in different levels of the system. Some more recent uses are highlight in green. The rows represent the different types of use, the columns the different foci of assessment of performance. These are not discrete categories or cells, and limited time and space precludes a more complete description of these uses. But, this display does provide a framework for identifying extant and potential measures of performance that are designed to meet specific purposes. The succeeding tables will begin to sketch out some of these measures.

Uses for students. We can look at how data has been used for students by working our way down the column titled "Students"). Teachers have historically used data on student performance to make instructional decisions about individual students, and teachers and schools use data to place students into instructional groups (within and across classrooms); programs (e.g., remedial, advanced; special education; ESL, etc.); and determine the kinds of services students might need (**diagnostic** purposes). Teachers and schools use data to determine whether students have successfully completed courses and have the knowledge to pass on to the next grade or graduate from high school (**evaluation)**.

School districts and states have also established criteria for promotion and high school graduation (**external accountability**). A few states and school districts use standardized tests results as one factor in making promotion decisions for elementary and middle school students. Most states require students to complete minimum coursework to graduate, and 24 states currently require students to pass a state test to graduate from high school (2 more states will come on line in 2012).

Uses for teachers. Teachers also use student data at the classroom level to plan lessons and to align curriculum and instruction with standards and tested skills and content (**inform teaching and learning)**. Schools and school districts (and now states) have developed measures of teacher performance to identify needs for teacher professional development and support (**diagnosis)**, and for evaluating teachers for retention and tenure. A few states and school districts have begun to use student performance data in pay for performance programs, and a few districts have sought to publicly report teacher performance measures. And policymakers are considering how to use teacher performance data to assess the effectiveness of teacher preparation programs.

Uses for schools. Measures of student and teacher performance may also be used diagnostically at the school level, to identify areas for technical assistance to schools, and to direct or re-direct resources to and within schools. These data may also be used by school staff to design and implement school improvement plans; by school staff to focus their attention on, and align curriculum to, specific skills and content; and/or to evaluate programs in schools. School districts (and now states) have measures of principal performance for evaluating principals for retention and in some cases for pay. Some school-level data are reported publicly and may be used by parents and community members to press educators to make changes, or by parents to choose schools. The most visible use of school-level data is for external accountability (rewards and sanctions) under both federal and state accountability policies.

Table 1: Uses of Multiple Measures

Uses	Student	Teacher	School
Diagnosis	Instructional decisions Placement Allocation of educational services	Professional development and support	Resource allocation Technical assistance
Inform teaching/ learning		Focus, align, redirect content and instructional strategies	Focus on, align curriculum to skills/content School improvement planning
Evaluation	Certification of individual achievement	Teacher performance	Program evaluation Principal evaluation
Public Reporting		Teacher performance	School performance Parent or community action
External Accountability	Course credit Promotion High school graduation	Renewal Tenure Pay	School sanctions and rewards Principal renewal, pay

The next 3 tables present the array of measures that can be used to meet each purpose.

Multiple Measures for Students (Table 2)

Teachers have multiple ways of **diagnosing** student performance to make instructional decisions about individual students: observation of students in class, teacher-made assignments and tests, homework, end of unit tests and even benchmark assessments. Teachers and schools may use grades, standardized assessments and measures of behavior, as well as teacher evaluation of student work, to place students into instructional groups and programs.

Summative evaluations of a student's performance (promotion into the next grade, high school graduation) can be based on a student's grades, other teacher evaluation of the student's work, standardized tests and, for high school graduation, accumulation of course credits on the student's transcript. The relative weight given to these measures can be determined at the school level (generally true for promotion), or at the district and state level (generally true for high school graduation and, in some cases, for promotion).

A few states require schools to include student performance on an end-of-course exam in a student's course grade (e.g., North Carolina), and a few districts and states place primary emphasis on students' tested performance as a condition of promotion into specified grades. And as mentioned earlier, about half of the states require students to pass a state examination as a condition of graduating from high school.

Table 2: Multiple Measures: Students

Uses	**Uses**	**Measures**
Diagnosis	Instructional decisions Placement Allocation of educational services	Teacher evaluation of student work • Teacher observation • Teacher assignments • Homework • Teacher-made tests • End-of-unit tests Grades Benchmark tests (district/state) Annual tests (district/state) Behavioral measures (attendance, etc.)
Inform teaching/ learning		
Evaluation	Certification of individual achievement	Grades Teacher evaluation of student work Transcripts Annual tests (district/state) High school graduation tests ("Comprehensive"/End-of-course)
Public Reporting		
External Accountability	Course credit Promotion High school graduation	Grades Teacher evaluation of student work Attendance Transcripts Annual tests (district/state) High school graduation tests ("Comprehensive"/End-of-course)

Multiple Measures for Teachers (Table 3)

In addition to informing instructional decisions about individual students, teachers will look across the performance of students to make adjustments in their classroom-level instructional content and strategies (**informing teaching/learning**). They may use the same measures—observation, teacher-made assignments and tests, end-of-unit tests and standardized tests—but look for patterns across students in the class (e.g., understanding of a particular concept). Principals and districts have rubrics for observing teacher practice for the purpose of **diagnosis and evaluation** and, looking as well at student behavior and perhaps student grades, have historically rated teachers as satisfactory or unsatisfactory.

There have been three recent changes in the ways that districts and states measure teacher performance, however. The first is the Inclusion of student tested achievement, generally from annual tests. Student achievement can be measured as a percent of the class meeting a specified benchmark, and/or an aggregation of growth from point 1 to point 2. The second change is the creation of new rating categories (e.g., 4: Highly Effective; 3: Effective; 2: Partially Effective; 1: Ineffective) that combine measures of practice and student achievement. The third change is the development of teacher observation rubrics by states.

Also new is **public reporting** of teacher performance measures and use of these new measures/ratings in renewal, tenure and pay decisions (**accountability**).

Table 3: Multiple Measures: Teachers

Uses	Uses	Measures
Diagnosis	Professional development and support	Observation of practice Student grades Student tested achievement • Benchmark tests (district/state) • Annual tests (district/state) Student behavior (attendance, etc.)
Inform teaching/ learning	Focus, align, redirect content and instructional strategies	At classroom level: Student work Teacher-made tests End-of-unit tests Benchmark tests (district/state) Annual tests (district/state)
Evaluation	Teacher performance	Observation of practice • District rubrics • State rubrics Student tested achievement • Annual tests (district/state) • Status/growth
Public Reporting	Teacher performance	Rating of teacher performance (combination of measures of practice, student tested achievement) Student tested achievement
External Accountability	Renewal Tenure Pay	Rating of teacher performance (combination of measures of practice, student tested achievement) Student tested achievement

Multiple Measures for Schools (Table 4)

Just as teachers look across the performance of students to make adjustments in their classroom-level instructional content and strategies, groups of teachers (by grade span, vertical) and school leaders may use some of the same data—teacher evaluation of student work, standardized tests—to look for patterns across classes in the school to **inform teaching/learning** at the school level (e.g., alignment with curriculum, content standards). This information, as well as data on teacher practice, can be used to design and evaluate the implementation of school improvement plans. Data on student performance and behavior and on teacher practice may be aggregated to the school level (by grade, subject, etc.) for purposes of **identifying technical assistance needs** to groups of teachers or for allocating or re-allocating resources (time; special programs for students).

Multiple measures can be used to **evaluate** a new program (e.g., reading curriculum) and a school's principal. Evaluation of a program might require information on how teachers implemented it through examples of teacher practice and student work, and the effects of the program on student performance and, if relevant, behavior. Districts have criteria for evaluating principals, but as with teachers, these criteria have expanded recently to include measures of student achievement.

State and federal accountability systems have required **public reporting** of several aspects of schools since the mid-1990s. While initially focused on tested student achievement and attendance, the range of elements has expanded to include measures of school climate and may, in the future, incorporate measures of teacher and principal performance. Finally, schools are held **accountable** for the level and growth of student achievement, student attendance and graduation rates, and in some districts, for school climate. For example, 15% of an elementary or middle school's grade in NYC's accountability system is based on the results of a school environment survey administered yearly to parents, teachers, and middle and high school students. Again, school accountability may expand to include measures of teacher and principal performance.

Table 4: Multiple Measures: Schools

Uses	Uses	Measures
Diagnosis	Resource allocation Technical assistance	Observation of teacher practice Evaluation of student work Student tested achievement • Benchmark tests (district/state) • Annual tests (district/state) Student behavior (attendance, etc.) School climate (student, teacher, parent surveys)
Inform teaching/ learning	Focus on, align curriculum to skills/content School improvement planning	Across classrooms: Observation of teacher practice Evaluation of student work Benchmark tests (district/state) Annual tests (district/state)
Evaluation	Program evaluation Principal evaluation	Observation of teacher practice Student tested achievement • School-designed tests • Benchmark tests (district/state) • Annual tests (district/state) • Status/growth Student behavior (attendance, etc.) Observation of principal practice • District rubrics • State rubrics
Public Reporting	School performance Parent or community action	Student tested achievement; graduation rates Student attendance School climate (student, teacher, parent surveys)
External Accountability	Principal: Renewal, pay School: rewards and sanctions	Rating of teachers, principal Student tested achievement; graduation rates School climate measures (student, teacher, parent surveys)

Issues in Using Multiple Measures (Figure 5)

Figure 5 arrays the measures from Tables 2-4 by use and level to show whether and to what extent the same measures are used for different purposes within any level; and are used for the same purpose across levels. For example, teacher evaluation of student work (which can include observation of students in class, teacher-made assignments and tests, homework, and end of unit tests) can be used at the student level for diagnosis, evaluation and external accountability (as input into promotion and, in some cases, high school graduation). These same measures may be used across levels for diagnosis of individual students, informing teaching and learning across classrooms, and for allocating resources within schools. Similarly, observations of teacher practice may be used at the teacher level for identifying areas of support and for evaluation of teachers for reporting and external accountability, and across levels for identifying technical assistance needs at the school level, evaluating new instructional programs, and for evaluating and rating schools. Table 5 identifies other measures (grades, standardized tests, student behavior, etc.) that are used in multiple ways within and across levels of the education system.

Having similar measures used for multiple purposes raises a number of issues. First, different uses (e.g., continuum from diagnosis to external accountability) embody different technical requirements for measures and how they are combined (e.g., standardization, reliability, validity). Take, for example, the use of measures for student diagnosis and teaching/learning. A continuous improvement model assumes that teachers will gather information about student learning on an ongoing basis and will interpret and use that evidence in ways that improve instruction. Thus measures for instructional purposes need to be timely and relevant to classroom instruction. Measures must generate continuous and actionable information. Ideally, measures are embedded in instruction with continuous feedback to the teacher. Flexibility is important; reliability and comparability across classes is not as critical at the classroom level. When measures are used for accountability, however, one needs standardization of content, administration and scoring because you are making comparisons, whether to a benchmark or to other students, groups or schools. Measures must be valid, reliable and fair.

However, although some of the structure and technical requirements differ by measure, measures used for instructional and accountability purposes do share some common requirements. Standards should reflect a model of learning/learning progressions. To generate valid information about student learning, all measures should be aligned to these standards and model of learning. And measures must include valid measures of student growth to guide instruction and to make accountability decisions (particularly for teachers).

Second, different uses call for data collection with different levels of frequency and at different points in time. For example, at the student level, on-going monitoring of student work for student diagnosis requires frequent assessment often during instruction, versus more periodic assessment for informing teaching/learning, versus summative measures for external accountability. Similarly, observation of teachers for purposes of assistance and/or identifying PD needs might focus on specific needs (e.g., classroom management, teaching a math curriculum) and focus on a specific period of time, while observations for formal evaluation might be more infrequent.

This leads to my third and final issue: to what extent can one measure be aggregated for use across purposes or levels of the system? To what extent do we need unique measures for a specific purpose at a specific level? Look again at the different ways in which **teacher evaluation of student work** may be used: (a) diagnosis for instructional purposes; (2) to inform evaluation of student achievement for e.g., promotion decisions; (3) to inform teaching/learning at classroom level and across classrooms; and (4) for diagnosis and evaluation at the school level. Some of these uses require aggregation of one student's work across a semester or year; others require aggregation of multiple students' work across a classroom or multiple classrooms. One answer has been calls for the development of "formative assessments" (for **diagnostic** uses) or "benchmark assessments" that can be aggregated across a semester or a year to comprise a summative measure for a student (**evaluative** use) and across students for **accountability** purposes.

Aggregation of any measure raises the same technical issues as with differences in use: as one aggregates across purposes or levels, we need standardization of content, administration, and scoring to ensure consistency and comparability. This raises the on-going tension between appropriate design of a measure for a specified use and what elements of that design that are lost due to the technical requirements of aggregation or other uses.

Table 5: Multiple Measures across Levels

Uses	Student	Teacher	School
Diagnosis	Teacher evaluation of student work Grades Benchmark tests (district/state) Annual tests (district/state) Student behavior (attendance, etc.)	Observation of practice Student grades Benchmark tests (district/state) Annual tests (district/state) Student behavior (attendance, etc.)	Observation of teacher practice Evaluation of student work Benchmark tests (district/state) Annual tests (district/state) Student behavior (attendance, etc.) School climate (student, teacher, parent surveys)
Inform teaching/ learning		At classroom level: Evaluation of student work Benchmark tests (district/state) Annual tests (district/state)	Across classrooms: Observation of teacher practice Evaluation of student work Benchmark tests (district/state) Annual tests (district/state)
Evaluation	Grades Teacher evaluation of student work Transcripts Annual tests (district/state) High school graduation tests ("Comprehensive"/End-of-course)	Observation of practice Annual tests (district/state) • Status/growth	Observation of teacher practice School-designed tests Benchmark tests (district/state) Annual tests (district/state) • Status/growth Student behavior (attendance, etc.) Observation of principal practice

Uses	Student	Teacher	School
Public Reporting		Student tested achievement Rating of teacher performance	Student tested achievement Rating of teacher performance Student attendance School climate measures (student, teacher, parent surveys)
External Accountability	Grades Teacher evaluation of student work Attendance Transcripts Annual tests (district/state) High school graduation tests ("Comprehensive"/End-of-course)	Rating of teacher performance Student tested achievement	Rating of teachers, principal Student tested achievement Student attendance Graduation rates School climate measures (student, teacher, parent surveys)