

## Measuring the Content of Instruction: Uses in Research and Practice

by Andrew C. Porter

This article describes tools for measuring the content of instruction, the content of instructional materials, and the alignment between these. Illustrative findings about the use of these tools are reported, and possible additional uses, both for research and practice, are discussed. The validity of data produced through use of these tools is found to be quite good. An agenda for future work is sketched—both for improvement of the quality and versatility of the tools and for use of the tools in research and practice.

The content of instruction plays a primary role in determining gains in student achievement (Gamoran, Porter, Smithson, & White, 1997; McKnight et al., 1987; Rowan, 1998; Schmidt, 1983a, 1983b; Sebring, 1987; Walberg & Shanahan, 1983). No one would be surprised by the statement that students are more likely to learn the content that they are taught. Certainly this is the assumption upon which much of today's standards-based education reform rests. Until recently, however, the content of instruction has largely been taken for granted—in education research and often in education practice (Porter, Floden, Freeman, Schmidt, & Schwille, 1988).

Teachers, as they interact with students, are the ultimate arbiters of what is taught (and how). They make decisions about how much time to allocate to a particular school subject, what topics to cover, when and in what order, to what standards of achievement, and to which students. Collectively, these decisions and their implementation define the content of instruction (Schwille et al., 1983). In making these decisions, teachers receive advice and support from a variety of sources. Some claim that teachers teach what is in the textbook; others claim teachers teach what is tested (Floden, Porter, Schmidt, Freeman, & Schwille, 1981). Policymakers hope teachers teach what is described in content standards. So, in addition to the content of instruction, we need to know the content of the various potential influences on teachers' content decisions; and of course, it is essential to know the content of the assessment instruments used to measure gains in student achievement.

A surprisingly long list of research questions and concerns about education practice motivates the research described here (Porter, 1991). Taxpayers and parents have the right to know what content students are taught in U.S. public schools. For example, when a student takes an algebra course, is algebra really the focus of instruction, and if so, what kinds of algebra are emphasized? If students are tracked, what are the differences in the

content of their instruction? Do students from low-income families and students of color experience the same opportunities to learn valued academic content as other students? Do teachers know enough about the content of each other's instruction to ensure that students experience a reasonable progression of content as they advance from grade to grade or course to course?

Knowing the content of instruction, educational materials, content standards, and professional development is key to monitoring the implementation and effects of education reform. On the one hand, there are questions about how well a particular reform of educational content is reflected in the policy environment in which teachers operate and in the supports that teachers receive to make their jobs manageable. On the other hand are questions about how well educational content reforms are being implemented. Is the content of instruction coming increasingly into alignment with the intentions of the reform?

Finally, the content of instruction is an essential variable in research on factors affecting student achievement. The content of instruction serves as an intervening variable in testing the effects of standards-based reform on student achievement gains (Gamoran et al., 1997). The content of instruction serves as a control variable in studies of the effects of pedagogical practices on student achievement gains. When the findings from the Third International Mathematics and Science Study (TIMSS) described the U.S. curriculum in math and science as "a mile wide and an inch deep," the conclusion was based on the content of U.S. math and science textbooks (Schmidt et al., 2001; Schmidt, McKnight, & Raizen, 1997). The inference was that student achievement in the United States is low, relative to other countries, because U.S. students do not study a focused and manageable set of content that can be mastered within the time constraints of schooling.

In what follows, I describe tools for measuring the content of instruction, the content of instructional materials (e.g., content standards, textbooks, and achievement tests), and the alignment between these (Porter & Smithson, 2001b). First, I describe these tools and illustrate their use. Next, I evaluate the tools' validity and value in education research and practice. Finally, I sketch an agenda for future work with the hope that others will join in this promising effort.

The tools described in what follows differ from other efforts to measure the content of instruction and alignment in two important ways. First, the tools allow independent and replicable descriptions of the content of instructional practice and instructional materials. A single language for measuring content ensures description at a consistent level of depth and specificity. Second, the language allows alignment to be measured across a large number of instructional materials and instructional practices.

Most current efforts to measure alignment start with one particular set of standards and then measure the extent to which assessments (e.g., Webb, 1997, 1999) or textbooks (e.g., American Association for the Advancement of Science, Project 2061, see <http://www.project2061.org/>) are aligned to that one set of standards. There is no quantitative index of degree of alignment or any ability to compare degree of alignment across states or among assessment standards and textbooks.

### Tools for Measuring Content and Alignment

#### *Creating a Uniform Language for Describing Content*

The past 25 years of research on teachers' content decision making have seen the development of the following three types of tools for measuring content and alignment:

- Surveys of teachers on the content of their instruction,
- Content analyses of instructional materials, and
- Alignment indices describing the degree of overlap in content between, for example, standards and assessment.

The central idea behind these tools is the development of a uniform language for describing content. It is this uniform language that makes it possible to build useful indices of alignment.

The language I have developed in my research consists of uniform descriptors of topics and categories of cognitive demand<sup>1</sup> that together can describe the content of instruction. For example, Table 1 illustrates a two-dimensional matrix that uses this language to describe mathematics content. The topic dimension lists some of the descriptors of mathematics topics: multiple-step equations; inequalities; linear equations; lines/slope and intercept; operations on polynomials; and quadratic equations. The cognitive demand dimension lists five descriptors of categories of cognitive demand:<sup>2</sup> (a) memorize; (b) perform procedures; (c) communicate understanding; (d) solve nonroutine problems; and (e) conjecture/generalize/prove (see Appendix for definitions of each category).<sup>3</sup>

Content of instruction is then described at the intersection between topics and cognitive demand, based on data gathered from teacher surveys. I ask teachers to indicate, for the past school year (a) the amount of time devoted to each topic (level of coverage)

and then, for each topic, (b) the relative emphasis given to each student expectation (category of cognitive demand) (see Figure 1).<sup>4</sup> I use a 4-point scale as follows:

- Level of coverage, (a) none/not covered; (b) slight coverage (less than one class or lesson); (c) moderate coverage (one to five classes or lessons); and (d) sustained coverage (more than five classes or lessons).
- Relative emphasis given to each category of cognitive demand, (a) no emphasis; (b) slight emphasis (less than 25% of time spent on this topic); (c) moderate emphasis (accounts for 25–33% of time spent on this topic); and (d) sustained emphasis (accounts for more than 33% of time spent on this topic).

These basic data are then transformed into proportions of total instructional time spent on each cell in the two-dimensional matrix defined by the language (e.g., Table 1). Across the cells in the content matrix, the proportions sum to 1 (Porter & Smithson, 2001a).

The two-dimensional language also can be used for content analyses of instructional materials. A key decision is the unit to be analyzed. For assessments, the unit is an item. Sometimes, a rule is made that no more than three cells in the content matrix can represent an item; sometimes no restrictions are placed on the number of cells representing a single item. When an item has more than one score point, the item is weighted according to its number of score points. For content standards, selecting the unit to analyze is more difficult. The most successful approach has been to pick the most specific version of the content standards and, within that, analyze the content of each objective, paragraph, or phrase.

#### *Creating Indices of Alignment*

Figure 2 illustrates the types of content alignment that might be described; each arrow in the figure depicts a particular alignment. Achievement can be more or less aligned to instruction, instruction to district standards and assessments, and district standards and assessments to state standards and assessments. These are all examples of vertical alignment. Horizontal alignment is a

**Table 1**  
**Content Matrix**

Topic	Category of cognitive demand				
	Memorize	Perform procedures	Communicate understanding	Solve nonroutine problems	Conjecture/generalize/prove
Multiple-step equations					
Inequalities					
Linear equations					
Lines/slope and intercept					
Operations on polynomials					
Quadratic equations					

Time on Topic	Elementary School Mathematics Topics	Expectations for Students in Mathematics				
		Memorize Facts/Definitions/Formulas	Perform Procedures	Demonstrate Understanding of Mathematical Ideas	Conjecture, Generalize, Prove	Solve Non-Routine Problems/Make Connections
<none>	1 Number sense / Properties / Relationships					
① ① ② ③	101 Place value	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③
① ① ② ③	102 Patterns	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③
① ① ② ③	103 Decimals	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③
① ① ② ③	104 Percent	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③
① ① ② ③	105 Real numbers	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③
① ① ② ③	106 Exponents, scientific notation	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③
① ① ② ③	107 Factors, multiples, divisibility	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③
① ① ② ③	108 Odds, evens, primes, composites	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③
① ① ② ③	109 Estimation	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③
① ① ② ③	110 Order of operations	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③
① ① ② ③	111 Relationships between operations	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③	① ① ② ③

FIGURE 1. Teacher survey.

measure of the consistency of standards and assessments within a district or state—that is, the degree to which these policy instruments deliver a coherent set of expectations to teachers.

Alignment is the core idea in systemic, standards-based reform (Smith & O’Day, 1991). An instructional system is to be driven by content standards, which are translated into assessments, curriculum materials, and professional development, which are all, in turn, tightly aligned to the content standards. The hypothesis is that a coherent message of desired content will influence teachers’ decisions about what to teach, and teachers’ decisions, in turn, will translate into their instructional practice and ultimately into student learning of the desired content.

But how best to measure the degree of alignment? As explained earlier, use of a uniform language for describing instruction, assessment, instructional materials, and content standards makes it

possible to build meaningful indices of alignment. Because content analyses and teacher surveys produce data of proportions in a content matrix, measuring alignment becomes a question of the extent to which the proportions in one content matrix (e.g., describing an assessment) match the proportions in another content matrix (e.g., describing standards) (see Figure 3).

I have experimented with different indices of alignment. A particularly promising index is

$$\text{Alignment Index} = 1 - \frac{\sum |X - Y|}{2},$$

where  $X$  denotes cell proportions in one matrix and  $Y$  denotes cell proportions in another matrix. The possible values of this index range from 0 to 1.0, with 1.0 indicating perfect alignment. Conceptually, the index is the sum of cell-by-cell intersects. For ex-

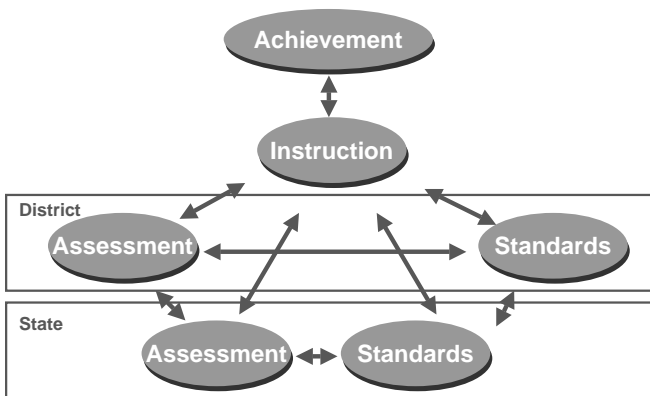


FIGURE 2. Vertical and horizontal alignment.

Topics	Cognitive Demand		
	Assessment	Standards	
	.3	0	.1
	0	.1	0
	0	.2	.1
	0	.1	.1
	.2	0	.1
	0	.2	0
	.1	.2	.1
	0	0	.1

$$\text{Alignment Index} = 1 - \frac{\sum |X - Y|}{2}$$

$X$ =Assessment Cell Proportions  
 $Y$ =Standards Cell Proportions

FIGURE 3. Example matrices to measure alignment.

ample, in Figure 3, the intersect in the (1,1) cell for assessment and standards is .2. Although one can say that the larger the value of the index, the better the alignment, there is still no easy way to think about how big the alignment index value must be to be considered “good.” The index does not have a straightforward interpretation like the proportion of common content between, say, standards and assessment. Nevertheless, the index does make it easy to see if content standards and student achievement tests are more aligned in one state than another.

Alternatively, one could calculate the correlation across cells between content proportions for two matrices. In theory, this cell correlation ranges from  $-1.0$  to  $1.0$ , but for alignment, the range is from  $0$  to  $1.0$ . In a study (Gamoran et al., 1997) using a data set derived from an end-of-year self-report survey of teachers’ instructional practices and the results of student achievement tests administered in the fall and spring, the alignment between test content and instructional content measured by the two indices correlated .86. At least for that data set, the two indices yielded quite similar values.

In my earliest work measuring alignment (Gamoran et al., 1997), I created what I called a “level index” and a “configuration index.” I defined *level* as the proportion of instructional content that was also tested and *configuration* as the degree to which the relative emphasis of tested content that was also taught matched the relative emphasis of tested content as a whole. I calculated configuration using the alignment index given earlier. An overall index of alignment was then formed by multiplying level by configuration. The level index correlated .62 with the cell correlation and .77 with the first alignment index. The configuration index correlated .72 with the cell correlation and .59 with the first alignment index. The first alignment index discussed above is the easiest to conceptualize and is the approach referred to in this article when the term *alignment index* is used.

### Use of the Tools to Study Alignment: Three Examples

#### *Measuring the Alignment of Assessments With Content Standards*

*Moving Standards to the Classroom: The Impact of Goals 2000 Systemic Reform on Instruction*, a U.S. Department of Education project conducted at the American Institutes for Research by Rebecca Herman and Laura Desimone, provides data for illustrating the alignment of content standards and assessments. This project used my content analysis tools to analyze the content of seventh-grade mathematics standards and tests in four states. They also analyzed the content of the National Council of Teachers of Mathematics (NCTM) standards (2000). Three raters independently analyzed state standards, and data averaged across the raters. For the tests and NCTM standards, there were two raters.

Table 2 presents the alignment of standards with assessments in the Goals 2000 study. To protect confidentiality, states are labeled B, D, E, and F. The table presents the results in a state-by-state matrix, where elements on the main diagonal represent the alignment of a state’s assessment with its own standards. The off-diagonal elements serve as a base for comparison. For states pursuing standards-based reform, where policy coherence is a goal, the main diagonal elements should exceed the off-diagonal

**Table 2**  
**Alignment of Assessments With Standards:**  
**Seventh-Grade Math—Goals 2000 Study**

Standard	Assessment			
	B	D	E	F
State				
B	<u>.37</u>	.39	.37	.45
D	.35	<u>.37</u>	.36	.40
E	.36	.33	<u>.43</u>	.31
F	.32	.35	.30	<u>.41</u>
NCTM	.34	.40	.33	.47

Note. Average within-state alignment = .40; average between-state alignment = .39; average state-test-to-NCTM alignment = .39.

elements (assuming each state has somewhat unique content standards). Surprisingly, this was not the case. The average within-state alignment was .40, and the average between-state alignment was .39. The within-state indices of alignment for States E and F were slightly higher than those for States B and D, probably because States E and F had grade level-specific standards, whereas States B and D had standards that covered a range of grade levels (including seventh grade). The average alignment of state tests to the NCTM standards was .39.

What can we conclude from these results? First and most important, because each standard and each assessment is mapped onto a common content language, it is possible to compare the alignment of assessment to standards within a state to the alignment between states and to NCTM. Most approaches to alignment of assessments to standards start with a particular state’s standards and ask to what extent the content in those standards is found on the test (Webb, 1997, 1999). Such analyses are unique to each state and do not allow between-state comparisons or comparisons between state and other professional standards. Second, we can conclude that the assessment of each state in the Goals 2000 study is no more aligned to its own standards than it is to the standards of the other states in the study or to those of NCTM. Perhaps state standards are not sufficiently specific to allow an assessment to be tightly aligned with them (more will be said about this later). Another possibility is that states have much more work to do to bring their assessments into alignment with their standards (Webb, 1999). This finding seems probable, and it is one about which U.S. Department of Education officials are expressing concern. Third, it could be that the content analyses on which the alignments in the Goals 2000 study are based are not sufficiently reliable to allow high degrees of alignment (more on this later). Finally, it must be recognized that tests are but a sample of items from a domain, whereas standards represent the domain. Thus, perfect alignment should not be expected.<sup>5</sup> Still, for states pursuing standards-based reform—and all states claim to be pursuing standards-based reform because it is a requirement for Title I funds—diagonal elements of alignment should exceed off-diagonal elements.

#### *Measuring the Alignment of Instruction With Assessments*

Data on instruction were not available from the Goals 2000 project, so I draw here on the Council of Chief State School Officers’

(CCSSO) State Collaborative on Assessment and Student Standards (SCASS) work, conducted jointly with the Wisconsin Center for Education Research. In the spring of 1999, 600 teachers from 20 schools across six states completed surveys to describe the content of their instruction in eighth-grade mathematics (Blank, Porter, & Smithson, 2001). Content analyses of these states' eighth-grade assessments were also completed. These data allowed investigation of the degree of alignment between instruction and assessments and offer a look at the extent to which teachers teach what is tested. In addition to the state assessments, we analyzed the content of the eighth-grade National Assessment of Educational Progress (NAEP) test.

The alignment results are reported in Table 3. Again, within-state alignment was hypothesized to exceed between-state alignment, but that was not the case. Within-state alignment was on average .22, and between-state alignment was on average .23. Surprisingly, instruction was more aligned to the eighth-grade NAEP (.39) than to state assessments. State O stands out as having a test that is not aligned to instruction in its state or to instruction in any other state (average alignment of instruction to State O's test was .05).

Although the alignment of instruction to assessments was not particularly high in any of these six states, it should be noted that the study did not include Texas or any other states with high stakes for students and schools attached to test results. Had Texas been one of our states, perhaps the alignment of its instruction with its assessment would have been much higher (McNeil, 2000). Still, the data do capture variance in alignment, as illustrated by the low alignment of instruction with State O's assessment, on the one hand, and the higher alignment of each state's instruction with NAEP, on the other. One cannot help but wonder whether the No Child Left Behind legislation will, over time, lead to increased alignment between instruction and NAEP, because NAEP is to be used to monitor the degree to which states are holding to high standards in their standards-based reforms.

#### *Measuring the Alignment of Instruction With Instruction*

The data from the SCASS study allow comparisons of instruction in one state to instruction in another state. Instruction was surveyed in a total of 10 states (the six for which there were content

analyses of state tests, plus four more). Table 4 presents the results, again in state-by-state matrix format. The off-diagonal elements describe the extent to which there is a national curriculum in eighth-grade mathematics. The average of off-diagonal elements is .69, indicating that instruction in one state is much more aligned with instruction in another state than it is, for example, with its own assessment or NAEP, or than a state's assessment is aligned with its content standards or the NCTM standards.

Again, the point of these displays is to illustrate use of these tools for measuring content and alignment. Nevertheless, the results may be of some substantive interest as well. Unfortunately, probability samples were not taken in each state; rather, the samples were of convenience. It remains to be seen whether the finding that eighth-grade mathematics instruction in one state is similar to that in another state would hold up when based on probability samples. I predict that it would.

#### **Use of the Tools to Map Content**

As we have seen, content descriptions using a uniform language make it possible to build effective indices of alignment. Although quantitative indicators of the degree of alignment among instruction, assessment, and content standards have their purposes, content descriptions are of use in their own right, permitting us to ask questions such as: When alignment is low, what is it about content that yields the low alignment? And what are the areas in which alignment exists?

I have experimented with a number of graphical displays to create powerful descriptions of the content emphasized (and not emphasized) in, for example, state content standards (Porter, 1998b). Figure 4 uses a topographical map display to indicate what content is emphasized by the content standards of States E and F in the Goals 2000 study and the NCTM content standards. These displays are at the coarse-grain level.<sup>6</sup> The first thing the maps indicate is that the standards exclude very little mathematical content. This finding is consistent with the TIMSS "mile wide and inch deep" conclusion. Nevertheless, there are some areas of emphasis. All three content standards emphasize number sense and numeration involving solving routine problems. State F and NCTM also emphasize students' communicating their understanding of number sense and numeration.

**Table 3**  
**Alignment of Instruction With Assessments:**  
**Eighth-Grade Math—SCASS Study**

Instruction	Assessment						
	H	J	K	L	E	O	NAEP
State	H	J	K	L	E	O	NAEP
H	<u>.35</u>	.22	.19	.28	.21	.04	.38
J	.34	<u>.21</u>	.18	.25	.20	.05	.38
K	.42	.28	<u>.21</u>	.29	.25	.05	.39
L	.36	.24	.19	<u>.29</u>	.22	.05	.40
E	.39	.26	.17	.26	<u>.24</u>	.04	.38
O	.35	.21	.16	.26	.20	<u>.05</u>	.38

Note. Average within-state alignment = .22; average between-state alignment = .23; average state-test-to-NAEP alignment = .39.

**Table 4**  
**Alignment of Instruction With Instruction:**  
**Eighth-Grade Math—SCASS Study**

State	H	J	K	L	E	O	G	I	M	N
H										
J	.73									
K	.59	.66								
L	.56	.64	.67							
E	.65	.71	.78	.70						
O	.71	.80	.63	.65	.70					
G	.71	.81	.66	.67	.71	.84				
I	.73	.82	.63	.66	.68	.79	.80			
M	.68	.77	.61	.62	.66	.73	.76	.79		
N	.62	.69	.58	.61	.62	.71	.70	.67	.65	

Note. Average alignment = .69.

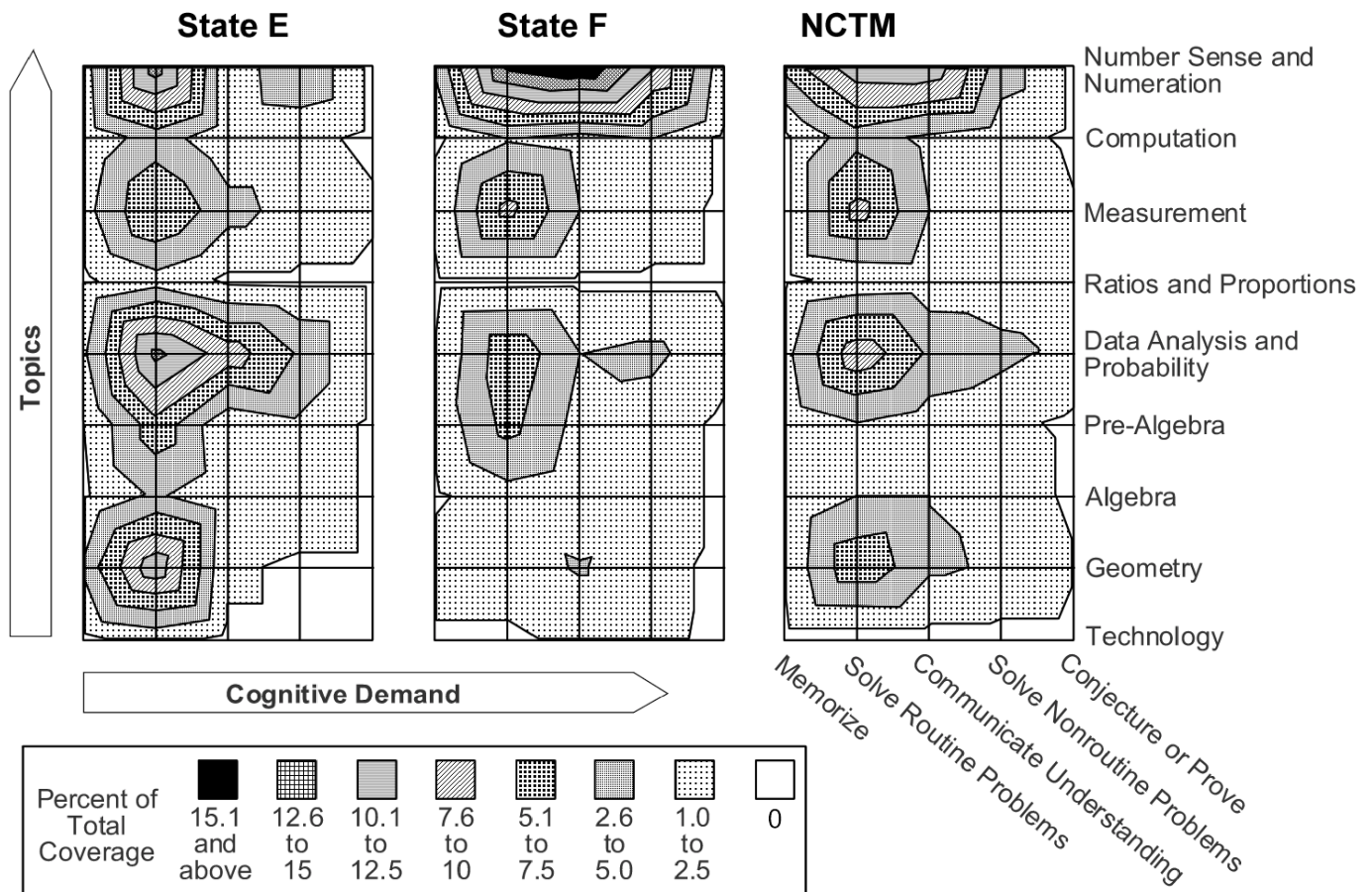


FIGURE 4. *Seventh-grade standards.*

State E is unique in its emphasis on solving routine problems in geometry and puts especially heavy emphasis on solving routine problems involving data analysis and probability; the NCTM standards put some emphasis on that, as well. Behind the coarse-grained graphic displays are finer grained descriptions. For example, Figure 5 allows a closer look at number sense and numeration. Here again, there are some similarities and some differences in content emphases. For example, NCTM standards are unique in their emphasis on communicating understanding of combinations and permutations.

There is some smoothing involved in creating the topographical maps. The shades between the intersections of topics and cognitive demand are where the smoothing is shown. Still, every data point in the content matrices of cell proportions is accurately depicted on the topographical maps. In contrast, bar graphs get quite busy, making it difficult to see the areas of relative emphasis and lack of emphasis (Porter, 1998b). Practitioners are quick to grasp how to interpret the topographical map displays and prefer them.

### The Quality of Data

The tools discussed in this article—surveys of teachers on the content of their instruction, content analyses of instructional materials, and indices of alignment—provide potentially powerful descriptions of content emphases and the degree of overlap in content between instruction, assessments, and content standards. But are the data valid?

Teacher surveys are valid only when teachers are willing to complete them. The surveys of instructional content are tedious to complete; teachers must carefully review a list of some 100 specific topics, indicating the degree of instructional emphasis given to each over the course of a school year. Then for all topics taught, teachers must estimate the distribution of instructional emphasis across the categories of cognitive demand. The task takes 45–90 minutes. Nevertheless, typical response rates are 75%, even for national probability samples of teachers (Garet, Porter, Desimone, Birman, & Yoon, 2001). More impressive, the percentage stayed at 75% even in a longitudinal study that required teachers to complete the survey once each year for 3 consecutive years (Porter, Garet, Desimone, Yoon, & Birman, 2000).

But completing the survey is only the first step. Teachers must also validly describe their instructional practices (Mullens & Gayler, 1999). Teachers may report what they think is appropriate, even if what they report does not accurately depict their own practice. This outcome is likely when data are to be used for high-stakes teacher evaluation purposes. Obviously, in a high-stakes context, a self-report survey is inappropriate. Teachers also may believe that their instructional content practices are different than they appear to third-party observers. Teachers may understand what content is wanted and believe they are teaching that content, when in fact they are not (Cohen, 1990). Finally, teacher respondents may not be clear on the terminology used in the survey instrument (Scherpenzeel & Saris, 1997; Sudman, Bradburn, & Schwarz, 1996).

## Number Sense and Numeration

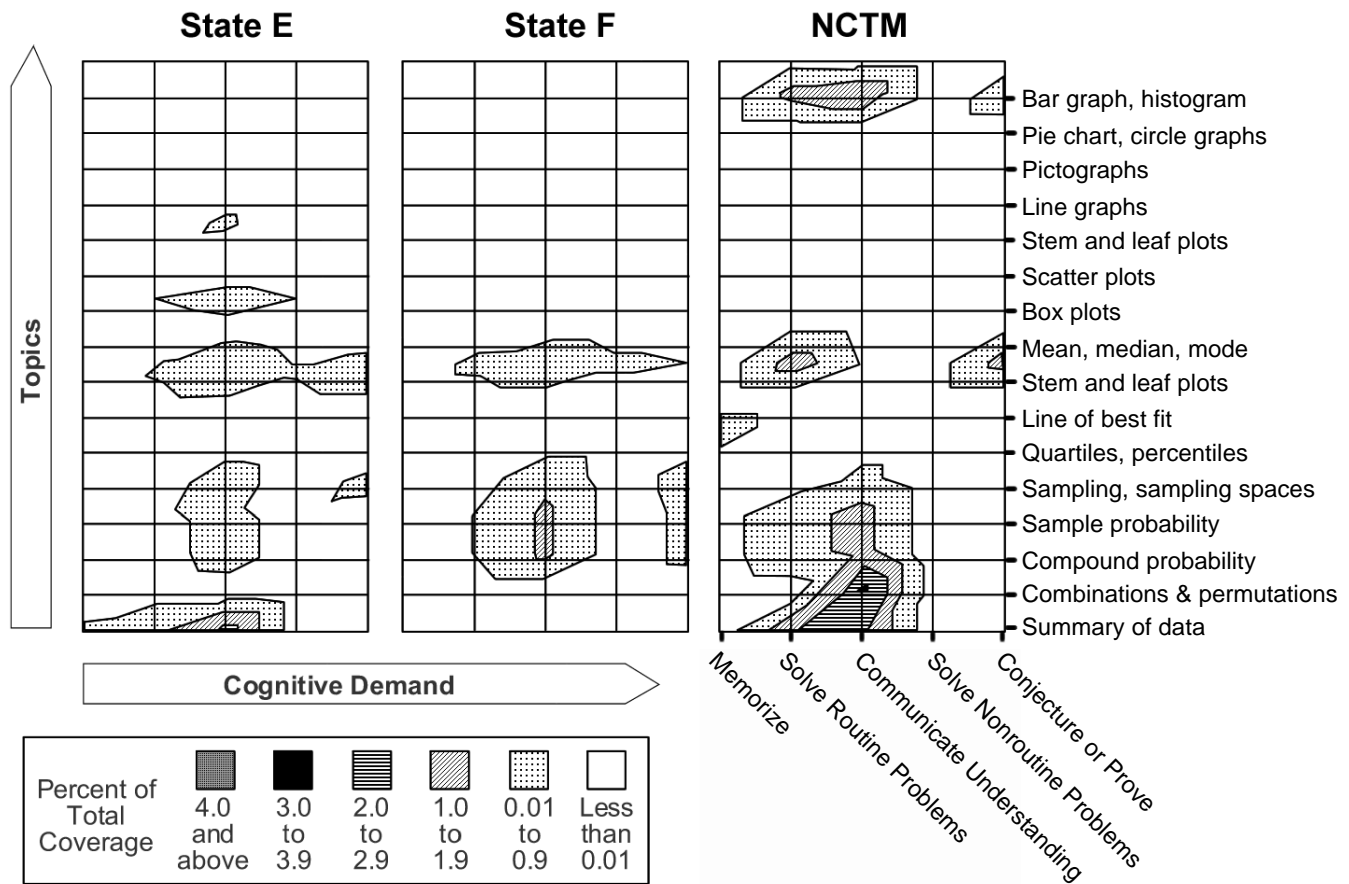


FIGURE 5. *Seventh-grade standards: Close view.*

Fortunately, a number of investigations of the validity of survey data for reporting instructional practice have been completed, especially in the context of mathematics instruction. Generally, these investigations find that survey data is excellent for describing quantity—for example, what content is taught and for how long—but not as good for describing quality—for example, how well particular content is taught (Burstein et al., 1995; Herman, Klein, & Abedi, 2000; Mayer, 1999; McCaffrey et al., 2001; Spillane & Zeuli, 1999). I investigated the validity of teacher self-report using the types of instruments described here with a sample of 62 teachers in 12 districts across six states (Porter, Kirst, Osthoff, Smithson, & Schneider, 1993), collecting end-of-semester survey data, daily teacher logs for a full school year, and observations of two lessons. Agreement between observations and logs on the days observed was high (correlations of .7 to .8). Agreement between daily logs aggregated to a full school year and end-of-semester surveys was quite good as well (correlations of .6 to .8 except in number, arithmetic, and measurement, where correlations were only .25 to .40).

A small digression is useful. End-of-year surveys have a number of strengths. They can be used with large national probability samples. They can capture what happens over an extended period, such as a full school year. They are easy to quantify, replicable, and inexpensive. When I have used daily logs to describe the content of instruction over a full school year, it has

been much more difficult to convince teachers to participate than when I have used only an end-of-year survey. If probability sample data on the content of instruction over an extended period is what is needed, then daily logs will not work. Observations are even more expensive and intrusive. They cannot be done on national probability samples, and they cannot be done daily for a full school year. When probability samples are not needed and when shorter periods of time are the target for description, then daily logs will provide better data than end-of-year surveys, and observations will provide even better data than daily logs. Unfortunately, for much of the work people wish to pursue, surveys are the only option. Fortunately, end-of-year surveys produce surprisingly valid data when the focus is on quantity, as it is here.

Admittedly, surveys have weaknesses. First, and perhaps most important, they are limited to what the developer of the survey decides to ask a priori. Logs suffer from the same problem, but observations are more flexible. Second, as was discussed above, end-of-year surveys are subject to self-report bias. Of course, so are daily logs, which are simply surveys conducted on a more frequent basis. Observations are not subject to self-report bias, but they can be subject to researcher bias. Finally, surveys are limited in the complexity of instructional practice they can capture.

There is yet another criterion for judging the validity of survey data on the content of instruction. If teacher survey data on

the content of instruction predict gains in student achievement, then surely the survey data must be valid. Admittedly, if survey data on the content of instruction do not predict achievement gains, the survey data could still be valid, but the pedagogical practices and the student effort were both so poor that learning did not occur. In a study of high school mathematics, 46 teachers across seven high schools in four districts in two states reported the content of their mathematics instruction through end-of-year surveys (Gamoran et al., 1997). Student achievement was tested in the fall and spring. For each teacher, the alignment between the content of instruction and the achievement test was calculated.<sup>7</sup> The alignment was .34 when calculated on student longitudinal data, and .31 when calculated on student cross-sectional data. These are strong correlations for predicting student achievement gains using instructional variables (Rowan, Correnti, & Miller, in press). The earlier described level-by-configuration index of alignment, although more complicated to describe and interpret, correlated .44 with cross-sectional data, but only .30 with longitudinal data.

The two-dimensional language for describing mathematics content allows descriptions at the intersection of topics and cognitive demand, but the language also can be reduced to topics only or cognitive demand only. The data set described in the preceding paragraph (Gamoran et al., 1997) was used to see if these simplified descriptions of content were strong predictors of student achievement gain. Unfortunately, the answer is no. Correlations for topics alone were approximately .15 and for cognitive demand alone, .07. Apparently, much would be lost by reducing the questionnaire to topics only (as was done in TIMSS) or to cognitive demand only.

Before leaving issues of validity, we must consider inter-rater agreement in content analyses. Generalizability theory is helpful. In the content analyses described earlier, the number of raters ranged from two to six. The reliability of average ratings across two raters was .70, and across four raters, .82. Surprisingly, inter-rater agreement was no higher for assessments than for content standards. I had imagined that the vagueness of content standards and the difficulty of knowing what unit to analyze would result in lower inter-rater agreement. That was not the case. These generalizability coefficients are for data reported at the specific topic-by-cognitive-demand level. Raters made judgments about which cells out of some 500 cells the content of a particular item or particular part of a content standard represented—not an easy task. Looking within the inter-rater agreement data, I found in each case one rater at odds with the others. Had there been a qualifying test for raters, or perhaps better training, the generalizability coefficients might well have been even higher.

### **Additional Uses of the Tools for Research and Practice**

The content analysis tools described here were developed in support of a 25-year program of research on teachers' decisions about what to teach in mathematics and science (Porter, 1998a). Although I have been discussing application of the tools to mathematics, these tools are just as developed for science. Some early work has been done in the areas of reading and language arts and social studies, as well. Over the past 25 years, my colleagues and I have made many uses of these tools for research purposes. In-

creasingly, however, the tools are being used for the improvement of practice.

In what follows, I offer some examples of how the tools can be used to describe (a) instructional practices, (b) instructional materials, and (c) alignment. In each case, I identify some uses for research and some uses for practice. I hope that readers will see potential uses for these tools in their own research or practice that will extend the work described here in exciting ways.

#### *Describing Instructional Practices*

First and most important to me, good measures of the content of instruction can serve to define the dependent variable in research on teacher decision making. The independent variables are the various messages that teachers receive about what should be taught. These messages come from the formal school hierarchy in the form of content standards, assessments, professional development, and curriculum materials. Messages about what to teach can come from individuals, as well—students and their parents, teachers in upper grades, the principal, or the district curriculum leader. And of course, messages can come from a teacher's own experience as a student, as a teacher in training (i.e., through preservice teacher education), and as a teacher (Porter et al., 1988). The research question asks to what extent do the content messages (independent variables) influence teachers' decisions about what to teach as reflected in their instructional practices (the dependent variables). For example, standards, assessments, and professional development might all seek to increase the teaching of probability and statistics in elementary schools.

Good measures of the content of instruction can also be used to describe the implemented curriculum or to measure the degree of implementation of a new curriculum. For example, one can determine whether there is grade-to-grade articulation in what is taught that makes sense from a student's developmental perspective (Porter, 1989). To take another example, when a new transition course called Math A was developed by California teachers, measures of the content of instruction provided evidence of the degree of implementation of this new course (Porter et al., 1993).

Measures of the content of instruction can be used to validate transcript studies. For example, when states increased high school graduation requirements in the late 1980s and early 1990s, transcript studies revealed a massive influx of new students into college preparatory math and science courses. Did these courses become watered down to accommodate the new and often academically weaker students? In a six state, 12 district, 18 high school study of math and science courses with the greatest increases in enrollment, surveys and observations were used to look beyond course titles at the actual content of instruction. Surprisingly, the content covered in these courses did not appear to have been watered down (Porter et al., 1993).

On the practice side of the equation, good measures of the content of instruction can provide the core of powerful professional development experiences. With National Science Foundation funding, my colleagues and I are conducting a randomized experiment on the effects of a professional development program on teachers' instructional practices. The professional development program begins with teachers' completing surveys to describe the content of their instruction. Graphic displays of results are returned to the teachers, who work in school teams to use the data to explore



(a) whether the content of their instruction is what they want it to be, (b) whether the differences in what teachers are teaching are appropriate, and (c) whether prerequisite courses are providing the content needed for effective grade-to-grade articulation.

### *Describing Instructional Materials*

Content analysis tools can facilitate research on the effects of textbooks on instruction. Curiously, relatively little work has been done in this area (Cronbach, 1955; Freeman & Porter, 1989). Perhaps the lack of research reflects an assumption that textbooks influence instruction. Yet in a study in which elementary school teachers kept daily logs of the content of their math instruction over a full school year and for which the content of their textbooks was analyzed, several distinctly different styles of textbook use were identified (Schmidt, Porter, Floden, Freeman, & Schwille, 1987). One style was to march through the textbook, starting at page 1 and going as far as time allowed. Even that style did not result in completion of the book nor did it involve using the conceptual development aspects of the book's lessons. Rather, virtually all of the time was allotted to the completion of exercises. Other teachers skipped chapters, such as measurement and geometry.

Content analyses of textbooks can be used to assess the breadth and the depth of the "intended" curriculum. The best example of this line of research is the analysis of textbooks done in TIMSS (Schmidt et al., 1997). The finding was that textbooks in the United States cover many more topics, in less depth, than those in other, higher achieving countries. Content analyses of instructional materials have many uses for the improvement of practice, as well. The best known use is for building student achievement tests. Any serious construction of a student achievement test begins by laying out test specifications that describe in detail the domain of content to be tested. The specific topic-by-cognitive-demand content matrix discussed earlier (Table 1) could be used to describe the domain. The items on the test would then be a representative sample of the domain. NAEP uses a less fine-grained two-dimensional content matrix (National Assessment Governing Board, 2001), as does Achieve.<sup>8</sup> The American Institutes for Research is using a two-dimensional content matrix in their test construction work for the state of Ohio.

Similarly, developers of content standards could use the content matrix. Surprisingly, though, I know of only one example in which such an approach has been taken. The El Paso Independent School District used a version of one of my content matrices to put together an Algebra 2 curriculum framework, essentially content standards for their Algebra 2 course (El Paso Collaborative for Academic Excellence, 2001).<sup>9</sup> The real power of using a content matrix as a tool for developing content standards is that it facilitates clarity in making hard choices. As was seen earlier, state content standards, and even NCTM standards, tend to include much more content than can be taught in depth. Use of the content matrix (and subsequent topographical maps displaying the content) can help build standards that clearly define not only what is to be taught but also what is not to be taught. For example, content standard developers could be limited to 100 points of emphasis to distribute across the content matrix. Such an arbitrary constraint might force difficult decisions about what content should be left out as well as what content should be emphasized.

### *Describing Alignment*

In research, indices of alignment between the content of instruction and a student achievement test can be used as a control variable for analyses of factors that explain variance in student achievement gains on the test. Thinking back to the days of process-product research on teaching, the paradigm was to administer a student achievement test in the fall and the spring, to observe pedagogical practices during the year, and to see which, if any, pedagogical practices predicted gains in student achievement (Brophy & Good, 1986). The content of instruction was neither measured nor controlled. Between-teacher differences in content covered were thrown into the error variance. The content of instruction is a powerful predictor of gains in student achievement; had content been used as a control variable in process-product research on teaching, undoubtedly that work would have been able to better identify pedagogical practices that contribute to achievement gains.

An index of alignment also can be used as a dependent variable, providing a quantitative test of the effects of standards-based reform. Such work requires repeated probability samples of the content of instruction (gained through teacher surveys)—say, once every 4 or 5 years. Content analyses of the content standards (and perhaps the assessments) are also required. State-by-state matrices of alignment between instruction and content standards are produced, one for each time point. If standards-based reform is having its intended effects, the main diagonal elements in the content matrices should become increasingly large and the off-diagonal elements increasingly small. This assumes that state content standards differ (and we have seen that they do). Of course, one should also have a theory to guide the work (Porter & Smithson, 2001a), and one potentially useful theory has been developed (Porter, 1994).

An index of alignment also can be used as a descriptive variable in assessing the coherence of a state's or district's curriculum policy system. At the heart of systemic reform is the concept of alignment. When a system is aligned, all the messages from the policy environment are consistent with each other, content standards drive the system, and assessments, materials, and professional development are all tightly aligned to the content standards. A policy instrument-by-policy instrument matrix of alignment could be generated. For standards-based content reform, the off-diagonal elements in this alignment matrix should all be large, indicating that the content messages of the various policy instruments are consistent.

In the field of practice, the previous Elementary and Secondary Education Act and the current No Child Left Behind Act require that states align their achievement tests to their content standards (Cowan & Manasevit, 2002). The U.S. Department of Education is rigorously monitoring the requirement. A state could use an alignment index, first, to build aligned assessments and, second, to make the case to the U.S. Department of Education (and to itself) that its tests are, in fact, well aligned to its content standards.<sup>10</sup> As yet one more example of the use of alignment for instructional improvement, consider again El Paso's Algebra 2 course. El Paso used the content matrix not only to establish its curriculum framework but also to build a coherent instructional system of materials and assessments aligned to that framework.

## An Agenda for Future Work

The content analysis tools discussed here are a work in progress. They can be extended and improved. What follows is a sketch of some areas in which further work is needed. The conceptual heart of the tools is a content language. A great deal of investigation and thought has gone into the development of the content languages in mathematics and science; and some work has been done in reading and language arts and social studies. I am sure the current versions of content languages can be improved. What is the right level of specificity in defining topics? What are the right number and types of cognitive demands? Is there yet a third or fourth dimension that could profitably be added to the language to capture more complex differences in the content of instruction? Are the labels in the language ones that have shared meaning among teachers? To this invitation to improve the content languages, I add two caveats. First, I am convinced that a perfect language is not possible. But second and more important, having one standard language per subject and grade level would have an enormous advantage, enabling cross-study comparisons and accumulation of knowledge. Work to develop a good content language for reading and language arts would be an especially useful contribution, if such a language were possible.

Closely related is the need for identifying contexts in which teachers' self-report of the content of their instruction is more or less accurate. As mentioned previously, I do not recommend the use of teacher self-report to provide data for high-stakes decisions. The temptation to look good would simply be too great for some. A more useful challenge is to determine the types of professional development that help teachers more accurately complete the survey questionnaires. Cognitive lab work might help to determine teachers' different interpretations of the topics and cognitive demands. Such information could be used to develop professional development experiences, from a seminar on accurate self-report to better instructions on the questionnaire. Helping teachers accurately describe the content of their instruction would almost certainly strengthen the profession. The same information might be used to build questionnaires that are more reliable and valid.

In my work on teacher content decision making, the most appropriate focus has seemed to be the content of instruction for a full school year (or at least a full course). End-of-year surveys are an efficient approach to collecting the information, but they do require teachers to accurately remember what was covered over an extended period. More frequent surveys are a solution to this problem (e.g., daily or weekly logs), but they are burdensome, making it impossible to do work using national probability samples. Another approach to reduce burden might be to take time samples. For example, teachers could report the content of their instruction for a single unit or week. I doubt that such time sampling would yield an accurate portrayal of what is taught over a full school year (Shavelson & Dempsey-Atwood, 1976), but I would welcome investigations of the possibility. The most convincing evidence of the validity of end-of-year surveys of instructional content is that alignment of survey-measured instruction to a student achievement test is a powerful predictor of gains on that test (Porter, 1998a). The finding is so important that it should be replicated.

Before doing the analyses, I had thought that efforts to improve the level and consistency of inter-rater agreement in content analyses were needed. But to my surprise, the generalizability coefficients are quite good, even for analysis of content standards that tend to be vague. Still, some work to improve inter-rater agreement would be welcome—in particular, work that identifies the odd rater marching to the beat of a different drum.

After trying several different indices of alignment, I have settled on the one reported in Tables 2–4 and defined in Figure 3. The index is a good predictor of gains in student achievement and is also the easiest to conceptualize. There may be other indices of alignment that are even better predictors of student achievement gains and are still conceptually simple. Development of indicators with these properties would be most welcome. At the same time, it would be good to know the distributional properties of any alignment index used. Knowing the distributional properties would allow hypothesis testing and building of confidence intervals to test—for example, whether the main diagonal elements in a state-by-state matrix of alignment between instruction and content standards are getting larger over time. At long last, there would be a quantitative test of the effects of systemic, standards-based reform.

Finally, the content analysis tools provide information that may be useful in developing new and more powerful programs for teacher professional development and for school data-based decision making. As described earlier, my colleagues and I with National Science Foundation support are using the tools to provide middle school math and science teachers feedback on the content of their instruction and the degree to which it aligns with assessments and standards.

## Conclusions

My purpose in this article has four parts: First, to illustrate the importance of the content of instruction as a variable in educational research and as a key component of efforts to improve the quality of instructional practice. Second, I offer tools for measuring the content of instruction, the content of instructional materials, and the alignment between these and illustrate the tools' validity and value. Third, I would like to stimulate thought about how these tools can be extended and improved. Fourth, in offering examples of how the tools have been used, I encourage thought about how the tools might be used in other educational researchers' work.

## NOTES

<sup>1</sup> Categories of cognitive demand distinguish what it is about a specific topic that a student is expected to know or be able to do.

<sup>2</sup> Over the years, researchers have tried using many different numbers and types of categories of cognitive demand, from as few as three to as many as 10 (Freeman et al., 1983; Porter et al., 1993). My current work uses the five listed here.

<sup>3</sup> In many ways, the two-dimensional language described here is like the two-dimensional language used to content-analyze textbooks in TIMSS (Schmidt & McKnight, 1995). Both have their conceptual origins in the work of the Content Determinants Group in the Institute for Research on Teaching at Michigan State University in the late 1970s to mid-1980s (Porter et al., 1988).

**Appendix**  
**Language Frequently Associated With Performance Goals**

A	B	C	D	E
Memorize facts, definitions, formulas	Perform procedures/solve routine problems	Communicate understanding of concepts	Solve nonroutine problems/make connections	Conjecture, generalize, prove
<ul style="list-style-type: none"> <li>• Recognize</li> <li>• Identify</li> <li>• Recall</li> <li>• Recite</li> <li>• Name</li> <li>• Tell</li> </ul>	<ul style="list-style-type: none"> <li>• Do computations</li> <li>• Make observations</li> <li>• Take measurements</li> <li>• Compare</li> <li>• Develop fluency</li> </ul>	<ul style="list-style-type: none"> <li>• Communicate mathematical ideas</li> <li>• Use representations to model mathematical ideas</li> <li>• Explain findings and results from statistical analyses</li> <li>• Explain reasoning</li> <li>• Describe</li> <li>• Select</li> </ul>	<ul style="list-style-type: none"> <li>• Apply and adapt a variety of appropriate strategies to solve nonroutine problems</li> <li>• Apply mathematics in context outside of mathematics</li> <li>• Analyze data, recognize patterns</li> <li>• Explore</li> <li>• Judge</li> </ul>	<ul style="list-style-type: none"> <li>• Complete proofs</li> <li>• Make and investigate mathematical conjectures</li> <li>• Infer from data and predict</li> <li>• Determine the truth of a mathematical pattern or proposition</li> </ul>

<sup>4</sup> As illustrated by Figure 1, teachers are asked first to review general areas of content—for example, *Number Sense/Properties/Relationships*—and then to review more specific topics under each general area—for example, *Place Value, Patterns, Decimals, Percent, Real Numbers*, etc., under *Number Sense/Properties/Relationships*. One important decision is how fine-grained to make the distinctions among topics. In my own research, I have typically used 100 or so specific topics grouped into 8 to 10 general areas.

<sup>5</sup> When a state has multiple forms of a test, the content of all forms should be analyzed to produce a single description of what is tested.

<sup>6</sup> Surface area charts can be created using a variety of charting software, including Excel.

<sup>7</sup> The achievement test was constructed from NAEP public release items and was designed to reflect a balanced curriculum across the categories of cognitive demand.

<sup>8</sup> Achieve is an independent, bipartisan, nonprofit organization created by governors and corporate leaders to help states and the private sector raise standards and performance in U.S. schools. Achieve was founded at the 1996 National Education Summit.

<sup>9</sup> Thanks to Lucy Michal and Susana Navarro for the opportunity to consult with them on this work and for keeping me up-to-date on their progress.

<sup>10</sup> The type of alignment index considered is one important piece of a more comprehensive view of alignment (Webb, 1999).

## REFERENCES

- Blank, R. K., Porter, A., & Smithson, S. (2001). *New tools for analyzing teaching, curriculum and standards in mathematics and science*. Report from Survey of Enacted Curriculum Project (National Science Foundation REC98-03080). Washington, DC: Council of Chief State School Officers.
- Brophy, J. E., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York: Macmillan.
- Burstein, L., McDonnell, L., Van Winkle, J., Ormseth, T., Mirocha, J., & Guiton, G. (1995). *Validating national curriculum indicators*. Santa Monica, CA: RAND.
- Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis, 12*(3), 311–330.
- Cowan, K. T., & Manasevit, L. M. (with C. J. Edwards). (2002). *The new Title I: Balancing flexibility with accountability*. Washington, DC: Thompson Publishing Group.
- Cronbach, L. J. (Ed.). (1955). *Text materials in modern education*. Urbana, IL: University of Illinois Press.
- El Paso Collaborator for Academic Excellence at the University of Texas at El Paso. (2001). *Algebra II curriculum framework*. El Paso, TX: Author.
- Floden, R. E., Porter, A. C., Schmidt, W. H., Freeman, D. J., & Schwille, J. R. (1981). Responses to curriculum pressures: A policy capturing study of teacher decisions about content. *Journal of Educational Psychology, 73*, 129–141.
- Freeman, D. J., Belli, G., Porter, A. C., Floden, R. E., Schmidt, W. H., & Schwille, J. R. (1983). The influence of different styles of textbook use on instructional validity of standardized tests. *Journal of Educational Measurement, 20*, 259–270.
- Freeman, D. J., & Porter, A. C. (1989). Do textbooks dictate the content of mathematics instruction in elementary schools? *American Educational Research Journal, 26*(3), 403–421.
- Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis, 19*(4), 325–338.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945.
- Herman, J. L., Klein, D. C. D., & Abedi, J. (2000). Assessing students' opportunity to learn: Teacher and student perspectives. *Educational Measurement: Issues and Practice, 19*(4), 16–24.
- Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*(1), 29–45.
- McCaffrey, D. F., Hamilton, L. S., Stecher, B. M., Klein, S. P., Bugliari, D., & Robyn, A. (2001). Interactions among instructional practices, curriculum, and student achievement: The case of standards-based high school mathematics. *Journal for Research in Mathematics Education, 22*(5), 493–517.

- McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J., et al. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective*. Champaign, IL: Stipes.
- McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- Mullens, J. E., & Gaylor, K. (1999). *Measuring classroom instructional processes: Using survey and case study field test results to improve item construction* (Working Paper No. 1999-08). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- National Assessment Governing Board. (2001). *Mathematics framework for the 2004 National Assessment of Educational Progress*. Washington, DC: Author.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Porter, A. C. (1989). A curriculum out of balance: The case of elementary school mathematics. *Educational Researcher*, 18(5), 9–15.
- Porter, A. C. (1991). Creating a system of school process indicators. *Educational Evaluation and Policy Analysis*, 13(1), 13–29.
- Porter, A. C. (1994). National standards and school improvement in the 1990s: Issues and promise. *American Journal of Education*, 102(4), 421–449.
- Porter, A. C. (1998a). The effects of upgrading policies on high school mathematics and science. In D. Ravitch (Ed.), *Brookings papers on education policy 1998* (pp. 123–172). Washington, DC: Brookings Institution Press.
- Porter, A. C. (1998b, October). *Curriculum reform and measuring what is taught: Measuring the quality of education processes*. Paper presented at the annual meeting of the Association for Public Policy Analysis and Management, New York.
- Porter, A., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1988). Content determinants in elementary school mathematics. In D. A. Grouws & T. J. Cooney (Eds.), *Perspectives on research on effective mathematics teaching* (pp. 96–113). Hillsdale, NJ: Lawrence Erlbaum Associates; Reston, VA: National Council of Teachers of Mathematics.
- Porter, A. C., Garet, M. S., Desimone, L., Yoon, K. S., & Birman, B. F. (2000). *Does professional development change teaching practice? Results from a three-year study* (Report to the U.S. Department of Education, Office of the Under Secretary, on Contract No. EA97001001 to the American Institutes for Research). Washington, DC: Pelavin Research Center.
- Porter, A. C., Kirst, M. W., Osthoff, E. J., Smithson, J. S., & Schneider, S. A. (1993). *Reform up close: An analysis of high school mathematics and science classrooms* (Final Report to the National Science Foundation on Grant No. SPA-8953446 to the Consortium for Policy Research in Education). Madison, WI: University of Wisconsin–Madison, Wisconsin Center for Education Research.
- Porter, A. C., & Smithson, J. L. (2001a). Are content standards being implemented in the classroom? A methodology and some tentative answers. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (100th Yearbook of the National Society for the Study of Education, part II, pp. 60–80). Chicago: National Society for the Study of Education; distributed by University of Chicago Press.
- Porter, A., & Smithson, J. (2001b). *Defining, developing, and using curriculum indicators*. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.
- Rowan, B. (1998). The task characteristics of teaching: Implications for the organizational design of schools. In R. Bernhardt, C. N. Hedley, G. Cattaro, & V. Svolopoulos (Eds.), *Curriculum leadership: Rethinking schools for the 21st century*. Cresskill, NJ: Hampton Press.
- Rowan, B., Correnti, R., & Miller, R. J. (in press). What large-scale survey research tells us about teacher effects on student achievement. *Teachers College Record*.
- Scherpenzeel, A., & Saris, W. E. (1997). The validity and reliability of survey questions: A meta-analysis of MTMM studies. *Sociological Methods and Research*, 25(3), 341–383.
- Schmidt, W. H. (1983a). High school course-taking: A study of variation. *Journal of Curriculum Studies*, 15(2), 167–182.
- Schmidt, W. H. (1983b). High school course-taking: Its relationship to achievement. *Journal of Curriculum Studies*, 15(3), 311–332.
- Schmidt, W. H., & McKnight, C. C. (1995). Surveying educational opportunity in mathematics and science: An international perspective. *Educational Evaluation and Policy Analysis*, 17(3), 337–353.
- Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. C., Wiley, D. E., Cogan, L. S., et al. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.
- Schmidt, W. H., McKnight, C. C., & Raizen, S. (1997). *A splintered vision: An investigation of U.S. science and mathematics education*. Dordrecht, NL: Kluwer.
- Schmidt, W. H., Porter, A. C., Floden, R. E., Freeman, D. T., & Schwille, J. R. (1987). Four patterns of teacher content decision making. *Journal of Curriculum Studies*, 19(5), 439–455.
- Schwille, J. R., Porter, A. C., Belli, G., Floden, R. E., Freeman, D. J., Knappen, L. B., et al. (1983). Teachers as policy brokers in the content of elementary school mathematics. In L. Shulman & G. Sykes (Eds.), *Handbook on teaching and policy* (pp. 370–391). New York: Longman.
- Sebring, P. A. (1987). Consequences of differential amounts of high school coursework: Will the new graduation requirements help? *Educational Evaluation and Policy Analysis*, 9(3), 257–273.
- Shavelson, R. J., & Dempsey-Atwood, N. (1976). Generalizability of measures of teaching behavior. *Review of Educational Research*, 46, 553–611.
- Smith, M. S., & O'Day, J. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing: The 1990 yearbook of the Politics of Education Association* (pp. 233–267). Bristol, PA: Taylor & Francis.
- Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, 21(1), 1–27.
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Walberg, H. J., & Shanahan, T. (1983). High school effects on individual students. *Educational Researcher*, 12(7), 4–9.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6). Madison, WI: University of Wisconsin–Madison, Wisconsin Center for Education Research.
- Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* (Research Monograph No. 18). Madison, WI: University of Wisconsin–Madison, National Institute for Science Education.

## AUTHOR

ANDREW C. PORTER is a professor at the University of Wisconsin–Madison and the Director of the Wisconsin Center for Education Research, 1025 West Johnson Street, Madison, WI 53706; andyp@education.wisc.edu. His research interests include psychometrics, research on teaching, and curriculum policy.

Manuscript received June 4, 2002

Revisions received June 27, 2002

Accepted July 9, 2002