DOCUMENT RESUME

ED 383 763 TM 023 306

AUTHOR	Kendall, John S.; Marzano, Robert J.
TITLE	A Report on the Findings of Phase I: The Identification and Articulation of Content Standards and Benchmarks.
INSTITUTION	Mid-Continent Regional Educational Lab., Aurora, CO.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Apr 95
CONTRACT	RP91002005
NOTE	40p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Databases; *Educational Change; Elementary Secondary Education; Evaluation Methods; *Identification; *Research Needs; School Districts; *School Restructuring; *Standards
IDENTIFIERS	*Benchmarking

ABSTRACT

The Mid-continent Regional Educational Laboratory (McREL), with funding from the Office of Educational Research and Improvement, undertook a project to address needs for school reform support that were identified as common among states in its region. In order to assist in a strategy for leveraging limited resources, McREL proposed to identify standards systematically to make them more accessible to states, districts, and local schools. From the inception of the project, it was clear that standards were described and articulated quite differently among the subject-area groups. Analysis of documents from the various groups has afforded the opportunity to propose or clarify distinctions that may be useful in identifying standards. Part of the process has been to recognize and point out the practical, and apparently unanticipated, consequences of some approaches that have been taken to identify content standards and benchmarks. Some considerations schools or districts may want to take into account are highlighted. To date, over 35 documents have been analyzed, and the project has produced a database of 338 standards and over 4,400 benchmarks. Appendix A lists standards and benchmarks, and Appendix B is a bibliography of 35 works cited in the project's database. (Contains 33 references, and 4 figures.) (Author/SLD)

A Report on the Findings of Phase I: The Identification and Articulation of Content Standards and Benchmarks

John S. Kendall Robert J. Marzano

Mid-continent Regional Educational Laboratory
2550 South Parker Road, Suite 500
Aurora, CO 80014

A paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1995.

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, Department of Education, under Contract Number RP91002005. The content of this publication does not necessarily reflect the views of OERI or any other agency of the U.S. Government.

Abstract

The Mid-continent Regional Educational Laboratory (McREL), with funding from the Office of Educational Research and Improvement, undertook a project to address those needs for school reform support that were found to be common among states in its region. In order to assist in a strategy for leveraging limited resources, McREL proposed to identify standards systematically so that states, districts, and local schools would find them more accessible and adaptable to their own plans for reform.

From the inception of the project, it became clear that standards were described and articulated quite differently among the subject-area groups. Over the past three years, the analysis of the documents from these and other organizations has afforded us the opportunity to propose or clarify distinctions that we believe useful for the identification of standards. Part of this process has been to recognize and point out the practical and apparently unanticipated consequences of some approaches that have been taken to identify content standards and benchmarks. As a result of this, we have also provided some considerations that schools or districts may want to take into account when they undertake to establish standards.

To date, over 35 documents have been analyzed and the project has produced a database of 338 standards and over 4400 benchmarks.

An earlier version of the database described in this report is currently available on the Internet via Mosaic or other World Wide Web browsers. The Universal Resource Locator (URL) is:

http://www.mcrel.org/standards-benchmarks/docs/std-ben.html

The standards and benchmarks are linked by hypertext, and the text is searchable.

A number of issues have been raised in the current dialogue on national standards and their role in school reform. Even across those groups who have accepted the challenge to develop standards, however, there is no clear consensus as to what form "standards" should take, or how they should be used. The result is that the character, scope, and level of detail provided in standards often vary significantly from one subject area to another. For schools, districts, and states who are currently struggling to develop standards, these differences quickly become problematic, since each subject-area competes for a common ground: the limited amount of time and resources in the school day. Unless standards and benchmarks are presented in a roughly equivalent and useable format, decisions regarding curriculum or assessment can quickly become problematic. For example, a standards-based system would be difficult for a school or district to implement if these standards could range from a paragraph describing a performance vignette, as is the case with the work done by the Standards Project for the Language Arts, to a description of specific components of knowledge and skills, as is the case with the mathematics standards developed by the National Council of Teachers of Mathematics. In addition, without a common format for standards, it is not likely that educators could recognize or take advantage of the possibilities for subject-area integration afforded by the commonalities that might otherwise be found across subject areas.

In the effort to develop a useful compendium of standards and benchmarks, it became apparent that a number of technical and conceptual differences among standards groups would have to be resolved to a certain degree. The authors here address six: 1) whether standards are for subject literacy, or subject expertise, 2) whether thinking and reasoning skills can be described independent of content, 3) whether standards should be formed as content or performance standards, 4) whether standards should be content or curriculum standards, 5) how benchmarks are defined, and 6) at what level of generality benchmarks and standards are stated. The model described here adopts a perspective on each.

The Literacy versus Expertise Issue

Some groups, such as the National Council of Teachers of Mathematics (NCTM), have developed standards using what might be called a "literacy" model. Such standards serve to ensure that students have a basic understanding of the fundamental knowledge and skills in mathematics that an educated, literate adult should know and be able to make use of. An indication that NCTM makes such a distinction can be seen in the standards the Council identifies separately for "the college-intending student." These standards appear to describe knowledge and skills important primarily for those in pursuit of advanced studies in math and science.

For example, in the document *Curriculum and Evaluation Standards for School Mathematics* (1989), NCTM singles out the following as applicable to students seeking to pursue mathematics at a post-secondary level:

- apply the sine and cosine functions to problem situations
- investigate limiting processes by examining infinite sequences, series and areas under a curve
- analyze graphs of polynomial, rational, radical and transcendental functions

A different view is available from the subject area of science. Project 2061 does not provide "expert" standards for students bound for advanced study. The title of the Project's work, *Benchmarks for Science Literacy* (1993), suggests that a distinction is to be made between knowledge that literate adults should possess and knowledge that is primarily of use to those who plan to do advance study in the field of science. This accords with another view of science literacy that "...*doing* science is clearly different from *using* science; scientific literacy concerns only the latter" (Hazen & Trefil, 1993). This does not mean, of course, that students should not engage in hands-on science; it merely suggests that there are distinctions that can be made between preparations for understanding science as an educated adult and doing basic science as an adult professional.

The differences between academic and literacy models presented in the various documents do not, on close analysis, constitute an insurmountable problem. At the literacy end of the continuum, standards might be described as the minimum requirements of knowledge and skill students should know and be able to do to function well as adults of the 21st century. At the "expertise" end of the continuum, standards are described in terms of the knowledge and skills that, once acquired, would render students "mini-experts" in every field. In fact, as currently articulated in the documents reviewed for this effort, both positions have strong tendencies toward the middle. That is, those documents that provide what might be characterized as literacy standards commonly include material that goes beyond minimum requirements for basic literacy within a domain. Additionally, those documents that are skewed toward the expertise position frequently are structured in such a way that the "expert-level" detail provided beneath a standard does not obscure the basic point of the standard itself, which focuses on information at a literacy level.

If one had to classify the model adopted in this project, it would be most accurately described as a literacy approach to content, in that it is believed that standards and benchmarks should be considered essential for all students, whether they enter the world of work directly from high school, or go on to higher education.

The Role of Thinking and Reasoning

Virtually all of the documents reviewed for this study either implicitly or explicitly acknowledged the importance of emphasizing thinking and reasoning in the articulation of standards. This is not surprising given the historical emphasis educators have placed on thinking and reasoning. Over 70 years ago, John Dewey wrote, "The sole direct path to enduring improvement in the methods of instruction and learning consists of centering upon the conditions which exact, promote and test thinking" (1916). Similarly, in 1961, the National Education Association identified the improvement of thinking and reasoning as central to American education:

> ...in the general area of the development of the ability to think, there is a field for

new research of the greatest importance. It is essential that those who have responsibility for management and policy determination in education commit themselves to expansion of such research and to the application of the fruits of this research. This is the context in which the significant answers to such issues as educational technology, length of the school year and content of teacher education must be sought and given. (Educational Policies Commission, 1961, pp. 14-15)

More recently, calls for the enhancement of thinking and reasoning in American education have come from the National Science Board Commission on Pre-college Education in Mathematics, Science and Technology (1983), the College Board (1987), the National Education Association (Futrell, 1987), and the American Federation of Teachers (1985).

Although there is agreement as to the importance of enhancing thinking and reasoning, there is not much agreement on the manner in which thinking and reasoning should be articulated in standards. There were three principal ways that thinking and reasoning skills were addressed in the documents reviewed for this project. One approach was to establish a set of standards on generic reasoning. For example, the document *Workplace Basics: The Essential Skills Employers Want* (Carnevale, Gainer & Meltzer, 1991) identifies Creative Thinking as one of the 16 skills that are important to the workplace. Thinking skills identified in this manner are stated as generic mental processes that cut across all content areas. A second approach can be found reflected in the NCTM's *Curriculum and Evaluation Standards*, which articulates a standard entitled Mathematical Reasoning. Within this category, those reasoning processes presumed to be specific to mathematics, but useful within the various subdisciplines of mathematics, are identified. Finally, the third perspective is exemplified by draft from the Geography Education Standards Project (1993), which describes performance standards. Here no set of standards nor any one specific standard addresses thinking and reasoning. Rather, the performances are described in such a way as to embed thinking and reasoning processes. To illustrate, consider the following standard which makes explicit the need both to evaluate information and solve problems:

- The student can evaluate the related merits of maps, globes and other geographic tools to solve problems. (p. 5)

A cursory review of the literature in cognitive psychology would seem to favor the latter two positions concerning an approach to thinking and reasoning skills. That is, strong arguments have been made against the isolation of thinking and reasoning skills (Glaser, 1984; Resnick, 1987). However, it is important to note that these arguments focus upon instruction, not upon the identificatión of standards. The case has been well articulated that thinking and reasoning should not be taught in isolation of specific content. Quite obviously, one cannot think about nothing. Thinking and reasoning processes and strategies must be employed with content, and to use any content other than that important to specific disciplines makes little sense. However, articulating standards is quite another matter. As described in this study, one of the primary purposes of standards is to provide educators with direction about the skills and abilities that should be the focus of instruction and assessment. Yet, if important thinking approaches are only found embedded in content, there can be no way to ensure that students have explored content in as many thoughtful ways as possible. To illustrate, consider the following performance standard from a draft of the national geography standards:

- The student can make and defend reasoned decisions on the location of a variety of activities within the home or community (p.11).

This performance standard describes one way in which a student might demonstrate knowledge of a content standard in geography. The important geographic knowledge within the performance, however, could be demonstrated in a variety of different ways. For example, the student could be asked to predict where types of activities might be located in a new community. In any case, once the geographic knowledge has been identified as important, it can be addressed in a number of different ways. But what of the ability to make and defend a decision based on knowledge of geography? If the ability is considered important enough that a student should be able to apply decision-making to issues in geography, then decision-making should be identified and addressed as systematically as the content, rather than as an incidental part of a performance standard in geography. Otherwise, whether a student uses decision making or not will be determined by the luck of the draw — only if he or she is asked to meet this particular performance standard or another performance standard that happens to require this skill. Clearly, such a hit-or-miss approach will characterize any effort that does not fully articulate and address the thinking and reasoning skills that should be brought to the study of content.

The second approach found within the various national reports also proved problematic. Many of the thinking and reasoning skills and abilities identified within those standards that purported to focus on content-specific skills and abilities were, in fact, quite general. For example, the NCTM standard of mathematical reasoning primarily specifies such general thinking and reasoning abilities as making conjectures, making inferences, and making corrections.

Finally, we found considerable redundancy in the thinking and reasoning skills and abilities implicitly and explicitly identified in the various documents. In effect, all the discipline areas not only emphasized thinking and reasoning skills and abilities, they generally identified the same skills and abilities. Given the intent of this project to capitalize on redundancies, and propelled by the problems described above, we have aggregated the various elements of thinking and reasoning into a dedicated set of standards.

Content or Performance Standards?

One of the significant distinctions within developing models of standards-based education is that between content and performance standards. Some theorists describe standards in terms of knowledge and skill that should be acquired. This is the content position. For example, Albert Shanker (1992), president of the American Federation of Teachers, defines a standard as "what we want students to know and be able to do as a result of their education" (p. S11). For Shanker, identifying a standard involves identifying specific information or skills that should be mastered to gain expertise in a given domain. Former Assistant Secretary of Education Diane Ravitch also describes standards from an information and skills perspective (Ravitch, 1992). Co ,versely, in his early writings on standards, Grant Wiggins, a nationally known expert on performance assessment, defined standards more in performance terms. For Wiggins, standards should be stated in terms of real-world, highly robust tasks that ideally elicit or require the important knowledge and skills within various content domains. The emphasis on performance as the critical feature of a standard is also shared by psychologist and researcher Richard Shavelson and his colleagues, who state that standards should be "based on students' performance of concrete, meaningful tasks" (Shavelson, Baxter & Pine, 1992, p. 22).

Since these earlier discussions, which seemed to pit one form against the other, clearer distinctions have been made and the two positions have been reconciled. The distinction between content standards and performance standards was perhaps formalized and legitimized in the 1993 report to the National Education Goals Panel by NESIC, the National Education Standards and Improvement Council. Commonly referred to as the "Malcom Report" in deference to Shirley M. Malcom, Chair of the Goals 3 and 4 Standards Review Technical Planning Group, the report makes a clear distinction between content standards and performance standards and establishes the validity of both:

> Content standards specify "what students should know and be able to do." They indicate the knowledge and skills - the ways of thinking, working, communicating, reasoning, and investigating, and the most important and enduring ideas, concepts, issues, dilemmas, and knowledge essential to the discipline - that should be taught and learned in school. (p. ii)
>
> Performance standards specify "how good is good enough." They relate to issues of assessment that gauge the degree to which content standards have been attained. While others use the term differently, in this report "performance standards" are not the skills, modes of reasoning, and habits mentioned above {in the description of content standards} that assessments attempt to measure.

> Instead, they are the indices of quality that specify how adept or competent a student demonstration must be. A performance standard indicates both the *nature of the evidence* (such as an essay, mathematical, proof, scientific experiment, project, exam, or combination of these) required to demonstrate that the content standard has been met and the *quality of student performance* that will be deemed acceptable (that merits a passing or "A" grade). (p. iii)

Performance standards, then, "contextualize" content standards by identifying the manner in which they must be demonstrated and the expected level of performance or understanding. In effect, performance standards are an interpretation of content standards. Both types of standards are, in fact, necessary. Their interdependence is explicitly referenced in the Malcom report: "The Technical Planning Group believes that performance standards are essential to gauging whether content standards are met" (p. iii). Additionally, reports from the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment (Shepard, 1993) and the National Council on Education Standards and Testing (NCEST, 1992) attest to the importance of both types of standards. In effect, then, content standards and performance standards are two interdependent and necessary components of an effective system of standards. As noted by Marc Tucker, Co-director of the New Standards Project:

> You can't assess kids' performance unless you give them the tasks, and you can't assess their degree of achievement unless they actually perform the tasks.
>
> But first you must be clear about what you want kids to know and be able to do, or what we call "content standards." Those content standards become the target for creating the assessment. (1992, p. S3)

Another reason for limiting the scope of this study to content standards is the necessary relationship content standards have with performance standards. That is, sound content standards are a necessary, but not sufficient condition for sound performance standards. Indeed, the NAEP efforts at setting performance standards have been criticized because they allegedly were not based on sound content standards. As Shepard notes:

> Current NAEP item pools, particularly at the advanced level, are not sufficiently congruent with emerging national content standards. Therefore, the achievement-level descriptions cannot adequately represent ideal future-oriented standards without departing from the assessment that the students actually took. In addition, some exemplar items were judged by content experts to be less than exemplary. They do not communicate subject-matter standards well. (Shepard, 1993, p. xiii)

In fact, Shepard implies that NAEP should curtail its efforts to set performance standards until content standards are well articulated: "Thus it only makes sense to wait until national content standards are available and then to follow a more coherent process for developing performance standards in conjunction with content standards" (p. xxv).

Again, given the developing nature of performance standards and their dependence on well-articulated content standards, we have chosen to focus our efforts on identifying content standards only.

Content Standards or Curriculum Standards?

A distinction that should be made in the types of standards various groups are identifying is that between content standards and curriculum standards. *Content* standards describe what a student should know and be able to do. *Curriculum* standards (sometimes referred to as program standards) are best described as descriptions of what should take place in the classroom; as such, they address instructional techniques, recommended activities, and various modes of presentation. The difference between a content and curriculum standard is illustrated by the following two statements from the National Council of Teachers of Mathematics (1989) framework. Within that document, both statements are presented as standards:

a) recognize when an estimate is appropriate

b) describe, model, draw, and classify shapes

Standard *a* describes a skill or ability a person might use solving a "day-to-day" or academic problem. For example, in day-to-day life, a person might use the skill of estimation to anticipate how much a proposed project might cost; or in a mathematics class, a student might use his estimation skills to determine that a problem can be solved without additional, unnecessary steps. In short, estimation is a skill that is commonly used or applied to solve common day-to-day problems or accomplish goals in academic settings. Standard *b*, "describe, model, draw, and classify shapes," does not share this characteristic. That is, it is difficult to imagine many situations that would demand the skill of being able to model, draw, or classify shapes, whether to solve an academic or day-to-day problem. Rather, this kind of activity seems appropriate as an instructional device to help students understand shapes or to provide a way for them to demonstrate their understanding of shapes. Standards like *a* above are referred to as content standards because they describe information or skill that is essential to the practice or application of a content domain. Standards like *b* are referred to as curriculum standards because they identify the instructional or curricular activities that might be used to help students develop skill and ability within a content domain. It might be said that curriculum standards describe the methods designed to help students achieve content standards.

This project has content standards as its focus. There are two principal reasons for this choice. First, content standards describe the goals for individual student achievement, whereas curriculum standards provide information that is ancillary toward reaching those goals. Second, curriculum standards, which usually focus on activities, projects or techniques, if interpreted rigidly, could leave teachers with little or no room for instructional diversity. That is, if teachers or administrators interpret curriculum standards as activities that must be performed, then teaching goals can too easily be equated with the activities performed, and actual student achievement loses its primary focus. When such a prescriptive attitude is taken toward activities, such activities often prove inefficient and time-consuming, leaving little room for experimentation and the refinement of new approaches to teaching.

Given the content (as opposed to curricular) orientation of this project, the standards identified will have the characteristics of content area knowledge. Specifically, the information that comprises standards identified within this project will generally fall into three broad categories representing three general types of knowledge. At a basic level, knowledge within any domain can be organized into the categories exemplified in Figure 1. The first column contains examples of knowledge that involves processes. These processes may or may not be performed in a linear fashion. For example, performing long division is a process: you perform one step, then another, and so on. Reading a map also involves certain steps, but these steps, unlike those in long division, do not have to be performed in any set order. You might read the name of the map first, then look at the legend, or you might just as effectively perform these steps in reverse order. Knowledge of this sort is usually called *procedural knowledge*. One might think of such knowledge as composed of the *skills and processes* important to a given content area.

Figure 1

Procedural	Declarative	Contextual
reading a map	democracy	know when to use a map instead of a globe
performing long division	a numerator	model numbers using number line
setting up an experiment	an amoeba	classify organisms
shooting a free throw	rules of basketball	know when to use man-to-man vs. zone coverage
editing an essay	conventions of punctuation	use appropriate tone and style for a selected audience

The examples in the second column do not involve a process or a set of steps. Acquiring this type of knowledge involves understanding the component parts. For example, knowledge of the concept of "democracy" includes understanding that decisions are made by the people, that each person has a single vote, that votes are weighted equally, and so on. This type of knowledge is commonly called *declarative knowledge.* One might think of such knowledge as composed of the *information* important to a given content area.

The last column contains items that are not simply declarative or procedural but specify knowledge in context. Column three contains examples of information and/or skills that have particular meaning because of the conditions that form part of their description. "To classify" is a skill; to understand the characteristics of organisms is declarative knowledge, or information; but knowledge of how to classify organisms is knowledge of a particular type: it requires understanding how particular characteristics establish relationships among organisms. Like the declarative/procedural distinction, this contextual knowledge is basic — a "piece" that cannot be further reduced without loss of important information. Also, like declarative and procedural knowledge, it reflects a kind of knowing that is primarily useful in the service of some larger goal. In the case of procedural knowledge, for example, the ability to read a map is not useful in itself, but does help when one needs to get somewhere. Similarly, declarative knowledge can prove useful when that knowledge helps in making inferences, decisions, and the like. Understanding the concept "democracy" is not *in itself* useful (excluding knowledge for its own sake, or for so-called "academic" exercises, such as asking whether the concept has been accurately defined). It is, however, useful in the service of some larger goal, as for example, determining whether activities within a country represent the democratic process. Similarly, contextual knowledge, of itself, has limited usefulness in that it does not describe a purposeful task, but its successful use may be essential for reaching a meaningful goal. A student may know the uses of a map and the uses of a globe, but in order to solve a distance measuring problem, the student should know the criteria for selecting between a map and a globe. This knowledge is different from simply knowing the uses of a map or of a globe (exclusively): it is knowledge about the conditions or context in which the use of one tool is more appropriate than the use of another.

How Benchmarks are Defined

Regardless of their position on standards, most groups acknowledge the need to identify expected or anticipated skill or understanding at various developmental levels. These statements of expected knowledge and skill are referred to as "benchmarks." To illustrate, consider the following content standards within science:

Understands basic concepts about the structure and properties of matter

At the 12th-grade level, the benchmark or expected level of understanding might be described in the following way:

- Knows that atoms interact with one another by transferring or sharing electrons that are furthest from the nucleus; these outer electrons govern the chemical properties of the element

At the 8th-grade level, the benchmark or expected level of understanding might be:

- Knows that atoms in solids are close together, and don't move about easily; in liquids, atoms are close together and stick to each other, but move about easily; atoms in gas are quite far apart and move about freely

Theoretically, these benchmarks, or subcomponents of a standard, could be identified at all grade levels. However, the trend seems to be toward developing benchmarks at a few key levels. For example, the National Assessment of Educational Progress (NAEP) identifies benchmarks at grades 4, 8, and 12. The American Association for the Advancement of Science (Project 2061) identifies benchmarks at grades 2, 5, 8, and 12. In this model, benchmarks identify expected understanding or skill at various grade levels, with a preference for articulating benchmarks at primary, upper elementary, middle, and high school within each standard. However, these levels are significantly different in some content areas, depending entirely on the availability of source materials.

Levels of Generality

The benchmark

The benchmark is the smallest unit of analysis for this study. As described above, it can be characterized as being declarative, procedural, or contextual in the type of knowledge it describes. The "size" of a benchmark is more problematic[1] and is best described, at this point in our study, in practical rather than theoretical terms. A practical description begins from what appears to be common among the benchmarks that we have identified within the subject areas.

From our observations, a benchmark seems to have a lower and an upper limit. As to the lower limit, in no case does it appear to describe specifics of knowledge or specific skills that an average student could master quickly, assuming that the benchmark has been placed at the appropriate grade level. This lower limit means that a declarative benchmark would never be equivalent to a short list of facts, for example, nor would a two-step algorithm be identified as a procedural benchmark at the 4th-grade level. This provides a rough starting point for the lower level of a benchmark.

A useful reference point for a benchmark, particularly at the lower end of the interval, is the behavorial objective. A benchmark is "larger" than a behavioral objective. Mager (1962) described what came to be called a behavorial objective as consisting of three key elements: a target behavior, a description of conditions under which the behavior is demonstrated, and criteria for acceptable performance. By limiting the description of knowledge and skill to a behavior and

[1]For a discussion on how the developers of Project 2061 approached the problem of "grain size," see *Benchmarks for Science Literacy*, p. 314-315.

to the conditions under which that behavior is demonstrated, this approach necessarily required many, many thousands of behavioral objectives to describe the knowledge within a given content domain. Benchmarks, by contrast, do not describe the behavior of students who meet an objective, nor do they narrow the description of knowledge and skills to a particular set of conditions.

Thus, a single behavioral objective could not cover a benchmark, but a single benchmark could be the source of a number of instructional objectives. This characteristic of benchmarks, at least as they appear in the database resulting from this project, is in part explained by the fact that the articulation of standards and benchmarks is not an attempt to organize learning or learning activities within a model for instruction. Rather, this approach uses a cognitive theory of knowledge types to assist in the analysis and identification of knowledge and skills.[2] At the lower limit, then, individual benchmarks do not prescribe instructional objectives. That is, as said in a report from the National Academy of Education Panel (Shepard, 1993) of the NCTM standards, they "do not delineate specific instructional activities, [but] they do set the direction for what should be taught" (p. 3).

In summary, a benchmark can be described as an "interval" of levels of generality in the description of knowledge and skills. In this section, we have attempted to describe some of the characteristics of the lower end of that interval. Benchmarks do not describe trivial or "easy" knowledge and skills for the developmental level at which they are found. They are not descriptions of knowledge and skill that have been narrowed through behavioral objectives or by a translation into an instructional activity.

---

[2]This process has been applied to documents, however, that have been developed by educators with understanding or belief about knowledge structures within their subject areas as well as what research says about the proper sequencing for the development of particular knowledge and skills. Clearly, then, the documents we have analyzed could well reflect the influence of certain theories of learning or theories of instruction. When this information (e.g., the sequence for learning about computation across K-12) is preserved in this study, it is better understood as a useful "side-effect" of our method, not a result of it.

The standard

Where the lower bounds of a benchmark have some identifiable characteristics, the characteristics of the upper bound are much more vague. That is, within this study it became difficult to determine the point at which the component of a standard seemed too broad in scope or too generally stated to be characterized as a benchmark. In fact, at the next broader level of generality, we found that, depending upon the document we analyzed, this level was either treated as a topic organizer or identified as a complete standard. The national history standards documents from NCHS were found to have at least four tiers of organization. In the design for the world history standards, for example, historical eras provided the most general structure. The level just beneath was identified as the standard level. Beneath the standard level, there was no detailed information, but three or four more specific statements were given (in our study, these were identified as standards), under which benchmark-level information was provided.

The subject area of science offers a convenient example of the variance in approaches to levels of generality, since two organizations have recently put considerable effort into the development of science standards, each using a different organizational scheme. Project 2061's *Benchmarks for Science Literacy* (1993) articulates most standards (termed "Literacy Goals,") across K-12. In practice, this means that a standard is described at a level that is broad enough to be articulated with benchmarks at each of four developmental levels (K-2, 3-5, 6-8, and 9-12). For example, one standard, or literacy goal, is on "the structure of matter." This idea is expressed at the earliest developmental level in terms such as the following:

By the end of 2nd grade, students should know that:

- *Objects can be described in terms of the materials they are made of (clay, cloth, paper, etc.) and their physical properties (color, size, shape, weight, texture, flexibility, etc.). (p. 76)*

At the upper level, 9-12, a sample benchmark under the same overarching idea is:

By the end of 12th grade, students should know that:

- *The configuration of atoms in a molecule determines the molecule's properties. Shapes are particularly important in how large molecules interact with others. (p. 78)*

Contrasting material comes from the National Committee on Science Education Standards and Assessment (NCSESA), which has been funded by the Department of Education to develop standards for science. If we search for an idea similar to the one found at the early grades in the *Benchmarks*, we find it in the following, which is identified as a content standard:

As a result of the activities in grades K-4, all students should develop an understanding of:

- *Properties of objects and materials*
- *Position and motion of objects*
- *Light, heat, electricity, and magnetism*

Concepts related to these topics, or subcomponents, are elaborated under a heading "fundamental ideas that underlie this standard." At that level, the following description is found for "Properties of objects and materials":

- *Objects have many observable properties, including size, weight, shape, color, temperature, and the ability to react with other substances (p. V-25).*

This demonstrates a dramatically different way of organizing very similar information. In this document, the standard has several organizing topics, each of which is defined at a greater level of detail. These details describe knowledge and skills at about the same level as found in the benchmarks from Project 2061's *Benchmarks*. The benchmark information differs essentially in two ways: in the NCSESA document, benchmarks appear in a standard that is complete at grade level, rather than articulated across grades, and these benchmarks also appear arranged under topic headings.

Although the categories differ, the same or very similar material is covered. For example, the corollary to the 12th grade benchmark from Project 2061 on the structure of molecules (see example above) can be found in the NCSESA as part of a different standard, which has six organizing subcomponents (p. V-130), under one of which ("structure and properties of matter") the following information can be found:

- *The properties of compounds reflect the nature of the interactions among its molecules, which are determined by the structure of the molecule (the kinds of atoms and the distances and angles between them). (p. V-134, V-135)*

In short, NCSESA has determined that standards should be categories of information not broad enough so that they encompass a common set of information across K-12. This articulation does appear, however, at the next larger level of organization. That is, all the benchmark information presented in the examples above from the NCSESA draft appears organized under the rubric Physical Science.

In this study, wherever possible, we describe standards at a level of generality that is broad enough to allow the articulation of benchmarks across K-12. Sometimes this approach required the reorganization of material from the subject-area documents. However, this organization was considered advantageous in that it organized knowledge and skills systematically across subject areas without any apparent loss of critical information. In addition, as mentioned at the outset, this consistency of format provides a clearer system-wide picture for those who wish to integrate benchmarks from different subject areas, but who also need to keep track of how and what curriculum they have addressed.

Summary

Standards, as found in the documents analyzed for this study, appeared at different levels of organization and structure. Standards provide a way of organizing information, those benchmarks that identify important declarative, procedural, and contextual knowledge. This organization itself may provide information on how "pieces" of knowledge can be sequenced, logically or psychologically, for students' ease of learning. In this project, the standards we have identified reflect both the character of the draft materials available to us and the model we have developed for identifying knowledge. There are other ways that benchmarks can be grouped, however, and except for the caution that developmentally sequenced information should not be lost, there appears no compelling reason why districts or schools should not feel free to organize benchmarks in whatever way they find most useful.

Application of the model

The considerations just outlined formed the rationale for the method we applied. Although some variations exist in the manner in which standards from different domains were addressed, a general process was followed to identify the standards in this project.

Identify National Reports

In February of 1990, President Bush and state governors established a set of national educational goals. One of those was that by the year 2000, American students would demonstrate mastery over challenging subject matter in core subject areas. Congress has since defined the goal areas to include the domains of English, mathematics, science, foreign languages, civics and government, economics, arts, history, and geography. In addition, the set of goals states that all students should have access to physical education and health education to ensure they are healthy and fit. Given this national mandate for improved student performance in these areas, the most significant documents in the fields were identified; standards are being identified in each of them and will be provided in the final report of this project. As of this writing (March 1995), documents in the following areas were identified: science, mathematics, history, geography, reading and writing, the arts, and health education.

In addition to these areas, documents were also reviewed for the domain of the workplace. Workplace standards, as made clear from recent Skills Standards efforts, funded be the Departments of Education and Labor, were developed to meet the growing demand for a smoother transition from school to the workplace.

It is important to note that a number of reference and supporting documents were in draft form. A list of the documents consulted thus far can be found in Appendix B.[3]

Select Reference Documents

Since there was more than one document within many of the domains considered, a reference report was selected for each domain. Reference documents were selected based on their completeness, perceived acceptance by the subject discipline community, and compatibility with the perspective on standards and benchmarks taken in this report. Over the course of the project, the selected reference document for several subject areas has changed, owing to the availability of finished material from those groups who have been funded to develop national standards.

Identify Standards and Benchmarks

Once a reference document was selected, standards and their benchmarks were identified. This was done from both "top-down" and "bottom-up" perspectives. A top-down perspective was taken when a reference document contained explicit standards that were at a level of generality consistent with position on standards taken in this study. In such cases, the standard found in the reference document was accepted with minor modifications, or if rewritten, kept close to the original meaning. Benchmarks were then identified for each standard. Depending upon the character of the document, this process could entail the straightforward identification of explicitly stated benchmarks, or an analysis of the material to find information that was more implicit. An

[3]For a complete discussion of the reference and supporting documents by subject area, see Kendall and Marzano (1995).

analysis would be required, for example, if knowledge and skills were found embedded within the description of an instructional activity. In some cases, however, a reference document articulated standards at a different level of generality (too general or too specific) or in a different format (performance or curriculum standards as opposed to content standards). In such situations, implicit and explicit benchmark components (declarative, procedural, and contextual elements) were identified first. These were then organized into standards. In effect, such standards were designed from the bottom-up. It was not uncommon that the standards so designed were influenced by other documents in the same field. For example, in the area of health education, a report from the Joint Committee on Health Education Terminology (1991) was found to provide a comprehensive set of topics that proved useful for organizing standards.

Integrate Information From the Other Documents

When the analysis of the reference document was complete, information from the other significant documents was then integrated into the standards and benchmarks which had been identified. These supplementary documents served a number of purposes: 1) to help us verify our readings of the reference document when the sense was unclear, 2) to ascertain and report for the reader, via citations, whether the knowledge or skill identified from the reference document was also recognized as important by the authors of other significant documents in the same subject area, and 3) to help determine the breadth of content coverage by the reference document. On some occasions, the analysis of material from secondary documents within a domain illustrated a need to create new benchmarks that were not explicit or implicit in the reference document.

Organize Standards into Categories

Thus far, the project describes 338 standards and their related benchmarks. These standards have been organized into eleven major categories. In a number of cases, the organization was straightforward; for example, standards generated from and referenced to science documents were placed under the category of science. Such an approach was followed for the areas of mathematics, geography and history. For other categories, the bottom-up approach, which characterized the formation of standards from benchmarks, also was used to organize similar standards into larger areas. Such was the case for the standards in Thinking and Reasoning. The number of standards and benchmarks are found listed by categories in Appendix A.

Structure of the Database Developed from the Application of the Model

The analysis, to date, of 35 relevant documents using the process just described has resulted in a database that currently contains a total of 338 standards and 4453 benchmarks in eleven areas. Across standards and benchmarks there are over 6300 citations. In addition to the citations, each benchmark is classified as declarative, procedural, or contextual knowledge, and identified by the developmental level in which it was placed. A number of subject areas, such as science, math, geography, and health, have benchmarks organized into the following four levels:

Level I	=	K-2, or primary
Level II	=	3-5, or upper elementary
Level III	=	6-8, or middle school
Level IV	=	9-12, or high school

Other standards areas analyzed in this project, however, have different grade ranges assigned for Levels I through IV. This difference of range is the result of the nature of the source material that was available for the area. For example, in the case of U.S. History, there are three levels, identified as:

Level II (Grades 5-6)
Level III (Grades 7-8)
Level IV (Grades 9-12)

Whereas in the standards for history at K-4, there are two levels, which are identified as:

Level I (Grades K-2)

Level II (Grades 3-4)

From this example it should be clear that level II is a relative description, defining grades 3-4 for history in the early grades, and grades 5-6 in the U.S. history standards.

Within each of the 11 categories, standards are numbered consecutively (the numbering sequence has no significance and was done for ease of reference). The benchmarks are grouped by standard and organized by level (I-IV). A set of codes, called a citation log, is associated with each benchmark in the database. To illustrate the structure of the database, a sample of the format used to publish the database on paper is provided in Figure 2. The citation log appears just above the benchmark, and flush right. The citation log provides the following information: the cognitive character of the benchmark (whether it describes declarative, procedural or contextual knowledge); a page number citation for each instance in which the information was found in reference and supporting documents; the nature of that citation (whether the information was found explicitly stated or could be implied from other statements); and finally, in the case of duplicates, where very similar benchmarks can be found within the same subject area.

Figure 2.

(GE,115)

5. Understands the concept of regions

Level I (Grades K-2)

BD (GE,115;EI,13;NI,35;TI,10;DI,4.1.2)

- Knows areas that can be classified as regions according to physical criteria (e.g., landform regions, soil regions, vegetation regions, climate regions, water basins) and human criteria (e.g., political regions, population regions, economic regions, language regions)

"Understands the concept of regions" appears as the fifth standard in the Geography section, and the benchmark shown is from Level I for grades K-2. Just above the benchmark, and flush right, is the abbreviation "BD," followed by the "citation log": (GE,115;EI,13;NI,35;TI,10;DI,4.1.2). A key such as that in Figure 3 is provided for each subject area.

Figure 3.

Codes (right side of page): BD = Benchmark, Declarative; BP = Benchmark, Procedural; BC = Benchmark, Contextual

1st letter of each code in parentheses	*2nd letter of code*	*Number*
G =National Geography Standards	E = Explicitly stated in document	Page number of cited document
E = Guidelines for Geographic Education	I = Implied in document	*or, for duplicates,*
N = NAEP: Item Specifications in Geography		Standard number & level of duplicate
T = K-6 Geography: Themes, Key Ideas		
D = Duplicated in another standard		

The key identifies "BD" as a benchmark that describes declarative knowledge. Within the parentheses that follow "BD," there are a number of documents cited, separated by semicolons. The first code, GE,115, indicates that the information described in the benchmark can be found explicitly stated (E) in the national geography standards (G) on page 115; the second citation, EI,13, indicates that the same information, although not explicitly stated, is implied in (or, can be inferred from) material on page 13 of the *Guidelines for Geographic Education*. Similarly, the same information can be inferred from two additional documents, the NAEP item specifications, and *K-6 Geography: Themes, Key Idea and Learning Opportunities* The last piece of information "DI,4.1.2" indicates that another benchmark contains very closely related information. In this case, that particular benchmark is under the standard number 4, at level 1, and is the second bulleted item.

Additionally, when the idea expressed at the standard level has been identified in supporting documents, that information is provided in parentheses, flush right, just above the standard statement. In the example above, the idea that students should have a general understanding of the concept of regions is found (GE,115) in the Geography Standards document on page 115.

The Implementation of Standards

The database of standards and benchmarks resulting from this project can be used in a number of ways. Before describing them, we must underscore the freedom that users of this database should feel to generate their own standards using ours as a reference point. This database was generated from basic assumptions and from a particular view of knowledge that have hopefully been well-defined for the reader. Other assumptions and other views of knowledge would, no doubt, have produced a very different articulation and organization of standards. Consequently, a school or district should feel free to extract benchmarks from our standards and organize them into other standards more consistent with their assumptions and perspective of knowledge.

Before using the database, a school or district should make some fundamental decisions and address some basic issues. In this section, we consider five important questions that a school or district should address before it adopts a standards-based approach to schooling.

1. In what format will benchmarks be articulated?

One fundamental question that should be addressed as early as possible is the format in which benchmarks will be articulated. There are two basic formats a school or district might use: 1) as lists of declarative, procedural and contextual knowledge, and 2) as knowledge application tasks. One approach to defining benchmarks is as leveled set of declarative, procedural and contextual knowledge. Clearly, this is the approach we have taken in this document. Consequently, a school or district that wishes to take this approach could simply select the standards from our listing and the benchmarks within those selected standards. However, a quite different approach is to state benchmarks as knowledge application tasks. If a school or district wished to state benchmarks as knowledge application tasks, it would use the declarative, procedural, and contextual benchmarks in our database to construct their more task or application-oriented benchmarks. For example, consider the following Level II benchmark from one of the standards in science:

- Knows that things that give off light often also give off heat
- Knows that mechanical and electrical machines give off heat
- Knows that heat can move from one object to another by conduction
- Knows that some materials conduct heat better than others; materials that don't conduct heat well can reduce heat loss
- Knows that electrical circuits require a complete loop through which the electrical

current can pass

Each of these might be used to construct explicit knowledge application tasks. For example, the first and fourth benchmarks might be translated into a knowledge application task in which students are presented with a specific situation where pairs of objects of different temperature — some warm and some cool — are placed in direct contact. Additionally, the paired objects would be made of materials of varying properties of conductivity. Some would be made of materials that are good conductors of heat; others would be made of materials that are poor conductors of heat. Students would be asked to hypothesize what changes in temperature would occur within each of the objects and explain the rationale underlying their answer. They might also be asked to test the accuracy of their prediction in light of the observed results and to describe alternative explanations for the results.

The benefit of constructing knowledge application tasks such as this is that they commonly include more than one benchmark. The proximity experiment, for example, involves two benchmarks from our original set. Additionally this knowledge application task can be used to determine student skill and ability in the thinking and reasoning standard, "Understands and applies basic principles of hypothesis testing and scientific inquiry":

- Makes and validates conjectures about outcomes of specific alternatives or events regarding an experiment

In short, articulating benchmarks as knowledge application tasks allows for the combining of information. As a way of using the elements of identified knowledge and skills, this combining can provide teachers with a powerful way of approaching content. However, articulating benchmarks as knowledge application tasks also presents some significant problems.

One of the most troublesome features of knowledge application tasks is that they are rarely transparent as to the knowledge and skill required for their completion. To illustrate, consider

Grant Wiggins' intriguing knowledge application task of calculating the cost of a shower (Wiggins, 1993, p.205). To ask students to determine the cost of a shower is an excellent real-world application of knowledge. However, if this task were used as an explicit benchmark, it is not immediately evident what declarative, procedural, or contextual knowledge the task is designed to assess. From surface appearance, it would seem that an understanding of the British Thermal Unit (BTU) is the declarative knowledge critical to the "shower" task. However, without explicit guidance one must make a calculated guess that this is the critical knowledge intended as the focus of the task. Conversely, if a knowledge of the BTU is explicitly stated as a benchmark, there are any number of tasks (including the shower task) that one could devise to confirm whether students grasp the central feature of the BTU.

Another problem inherent in stating benchmarks as knowledge application tasks is that they impose a rather rigid set of expectations on what will occur in classrooms. Given that benchmarks are written as knowledge application tasks, then all students must perform those tasks as evidence of their knowledge and skill in the benchmarks. In effect, teachers have no options as to how they will gain information about students' performance on the benchmarks when, in fact, many options might be available such as traditional forms of tests (e.g., essay tests, multiple choice, matching). Benchmarks as knowledge application tasks, then, have the same disadvantages as curriculum standards — they leave little room for divergence and experimentation in the classroom and mandate time-consuming activities.

2. How many standards and benchmarks will be articulated?

In all, this database, still to be updated, lists 194 standards and 2,787 benchmarks for implementation in K-12 schooling[4]. Clearly, a school or district could not expect a student to demonstrate competence in all of these (although they may be a part of instruction); sheer numbers would make such a system untenable. Given that there are 180 days in the school year and 13 years of schooling (assuming students go to kindergarten), there are only 2,340 school

[4]The total standards and benchmarks in the database (338 and 4453 respectively) differ from the total standards for implementation. See Appendix A.

days available to students. If all benchmarks in the project database were addressed, this would mean that students would have to learn and demonstrate mastery in one or more benchmarks every school day, or about six benchmarks every week.

Thus, a school or district will surely have to select from the standards and benchmarks available in the database if it wishes to construct a system in which students are to be held accountable for each benchmark. A reasonable number of benchmarks seems to be about 600, distributed in roughly the following way:

Level I:	K-2:	75
Level II:	3-5:	125
Level III:	6-8:	150
Level IV:	9-12:	250

Quite obviously, to implement this 600-benchmark cap, schools and districts would have to exclude quite a few of the benchmarks identified in this project.

3. Will all selected benchmarks be considered necessary to demonstrate competence in a standard?

One possible way to alleviate the problem of too many benchmarks is to consider benchmarks as exemplars rather than as necessary components of a standard. Using this option, students would be held accountable for demonstrating a mastery of a sample of the benchmarks within a level for a given standard as opposed to all the benchmarks within a given level. To illustrate, consider the benchmarks in Figure 4. for the science standard "Understands energy types, sources, and conversions, and their relationship to heat and temperature." A school or district that takes the "exemplar" approach to benchmarks would require students to demonstrate competence in a selected number of benchmarks per level. For example, a school or district might require students to demonstrate competence in two out of the three benchmarks for Level I; three out of five for Level II; five out of seven for Level III; and six out of eight for Level IV. This approach would

Figure 4

Level I (Grades K-2)

- Knows that the Sun applies heat and light to Earth
- Knows that heat can be produced in many ways (e.g., burning, rubbing, mixing chemicals)
- Knows that electricity in circuits can produce light, heat, sound and magnetic effects

Level II (Grades 3-5)

- Knows that things that give off light often also give off heat
- Knows that mechanical and electrical machines give off heat
- Knows that heat can move from one object to another by conduction
- Knows that some materials conduct heat better than others; materials that do not conduct heat well can reduce heat loss
- Knows that electrical circuits require a complete loop through which the electrical current can pass

Level III (Grades 6-8)

- Knows that energy comes in different forms, such as light, heat, chemical, nuclear, mechanical and electrical
- Understands that energy cannot be created or destroyed, but only changed from one form to another
- Knows that the Sun is a major source of energy for changes on the Earth's surface; the Sun's energy arrives as light with a range of wavelengths consisting mainly of visible light with significant amounts of infrared and ultraviolet radiation
- Knows that heat energy moves in predictable ways, flowing from warmer objects to cooler ones until both objects are at the same temperature
- Knows that heat can be transferred through materials by the collisions of atoms or across space by radiation; if the material is fluid, currents will be set up in it that aid the transfer of heat
- Knows that electrical circuits provide a means of converting electrical energy into heat, light, sound, chemical or other forms of energy
- Knows that in most chemical reactions, energy is released or added to the system in the form of heat, light, electrical or mechanical energy

Level IV (Grades 9-12)

- Knows that although energy can be transferred by collisions or waves and converted from one form to another, it can never by created or destroyed, so the total energy of the universe is constant
- Knows that all energy can be considered to be either kinetic energy (energy of motion), potential energy (depends on relative position), or energy contained by a field (electromagnetic waves)
- Knows that heat energy consists of random motion and the vibrations of atoms, molecules, and ions; the higher the temperature, the greater the atomic or molecular motion
- Knows that energy tends to move spontaneously from hotter to cooler objects by conduction, convection or radiation; similarly, any ordered state tends to spontaneously become less ordered over time
- Knows that the energy of waves (electromagnetic and material) can be changed into other forms of energy (e.g., chemical and electrical), just as other forms of energy (chemical and nuclear) can be transformed into wave energy
- Knows that some changes of atomic or molecular configuration require an input of energy, whereas others release energy
- Knows that each kind of atom or molecule can gain or lose energy only in particular discrete amounts and thus can absorb and emit light only at wavelengths corresponding to these amounts; these wavelengths can be used to identify the substance
- Knows that fission is the splitting of a large nucleus into smaller pieces, and fusion is the joining of two nuclei at extremely high temperature and pressure; nuclear reactions convert a fraction of the mass of interacting particles into energy

allow a school or district to meet a larger number of standards without exceeding the recommended limit of 600 benchmarks discussed in the preceding section. It would also allow for more flexibility within the classroom, in that individual teachers would have the option to use those benchmark components which they judged most applicable for their students. However, this approach also results in less continuity of coverage within a content domain since different teachers will no doubt select different benchmark exemplars to illustrate student competence within the levels for a given standard. It is also important to note that this approach may defeat the designed purposes of some well-articulated standards, such as those developed by Project 2061, where upper-level benchmarks are predicated under the assumption that students are familiar with a logically prior concept addressed at an earlier level. If teachers select without regard to articulation, some of the value of this approach may be lost.

4. Will student performance be reported using course grade or standards?

Currently, most schools and districts report student progress using appropriate grades for broad academic areas organized within courses. However, current research and theory indicate that courses of the same title do not necessarily cover the same content (Yoon, Burstein & Gold, undated). In other words, two courses of the same name do not necessarily cover the same declarative, procedural, and contextual knowledge. If a school or district wished to use traditional grades but implement a standards-oriented approach, it would ensure that the benchmarks that have been identified would be distributed systematically throughout the various courses within content areas, that is, specific benchmarks would be assigned to courses based upon the elements they cover. Any two courses with the same title would not only cover the same benchmarks but would place the same relative importance on the benchmarks they cover.

For example, assume that two courses of the same title were designed to cover the same seven benchmarks. The school or district could also determine which percentage of the grade each benchmark would command. In such a case, it might be determined that the first two benchmarks each accounted for 25% of the grade and the remaining five benchmarks each accounted for 10% of the grade. Clearly, this would provide more precision for course descriptions and show an equivalence between "identical" courses that is not often found today.

In summary, traditional grading practices and standards-based assessment are not incompatible. A school or district must simply distribute and weight the standards that have been identified across the various courses in a systematic, well-reasoned fashion.

The second reporting option a school or district might adopt is to report student progress by benchmarks. Rather than assign a single grade to a course, a teacher would report progress in some way for each benchmark covered in the course. In effect, for assessment purposes only, each benchmark component would be considered independent of the others covered within the course.

When this approach is taken, schools and districts commonly employ rubrics as opposed to grades. A rubric is a description of the levels of understanding or skill for a given benchmark. For example, below is a rubric for the Level II mathematics benchmark "Understands the basic role of place value":

4. Demonstrates a thorough understanding of the role and function of place value and provides insights that are not obvious when using the concept of place value.
3. Demonstrates a complete and accurate understanding of the role and function of place value as it relates to estimating or calculating addition, subtraction, multiplication and division.
2. Displays an incomplete understanding of the role and function of place value as it relates to estimating or calculating addition, subtraction, multiplication or division.
1. Has severe misconceptions about the role and function of place value as evidenced by severe place value errors in addition, subtraction, multiplication or division.

Commonly, one of the described levels within a rubric is designated as the targeted level of skill or knowledge. For example, a score of 3 in the reporting rubric above might be selected as the target standard for the Level II mathematics benchmark "Understands the basic role of place value."

Reporting out by benchmarks would, of course, require a record-keeping system that is far different from that currently used in most schools and districts. Each student's score on individual benchmarks would be recorded. Assuming the use of a four-point rubric, individual students would receive a score of 1 through 4 on each benchmark assessed within each standard. These scores could then be averaged to obtain an overall standard score at a given benchmark level.

5. Will all students be required to meet all standards?

A major decision facing a school or district who wishes to emphasize content area standards is whether students will be required to meet a targeted level of knowledge and skills. This approach is reminiscent of the mastery learning approach of the 1970s and early 1980s (see Levine & Associates, 1985) and the more recent outcomes-based approach or OBE approach (Spady, 1988). In the context of the reporting rubric described previously, a mastery or outcomes-based approach would mean that students would be required to receive a score of 3 on each benchmark. If a student did not meet the targeted level for a benchmark (i.e., did not obtain a score of 3 on the rubric), he or she would be provided with additional instructional opportunities until he or she could meet the required proficiency. Of course, such a system makes extreme demands on resources. In a traditional system, no extra resources need be used if a student does poorly in a course. In a mastery or OBE system, each student who does not meet a standard must be provided with whatever instructional and curriculum resources are necessary to ensure that the student meets the requirements. A variation in the theme of a comprehensive mastery or outcomes-based approach is to require that students meet the performance standards on *some*, but not all, benchmarks. Those benchmarks that are applied to all students would be considered a set of core requirements.

In summary, there are many important decisions a school or district must make regarding the implementation of a standards-oriented approach to schooling. In this section, we have discussed five of the decisions that deal with the nature and function of standards and benchmarks and the extent to which students will be held accountable for them.

APPENDIX A

The standards and benchmarks identified in the project's database

	Standards	Benchmarks
Science	18	321
Mathematics	9	294
History		
Historical Understanding	2	42
K-4 History	21	122
As Implemented[1]	±10	±61
U.S. History	91	950
As Implemented[1]	±30	±327
World History		
Core standards[2]	71	992
As Implemented[1]	±24	±330
Related standards[2]	38	480
As Implemented[1]	±13	±160
Geography	18	238
Connections	1	13
Dance	6	62
Music	7	80
Theatre	6	72
Visual Arts	5	42
Arts (total)	25	269
Language Arts	11	320
Health	10	126
Thinking and Reasoning	6	118
Working with Others	5	53
Self-regulation	6	59
Life Work	7	69

[1]The numbers shown for the history standards are not equivalent to numbers in other subject areas, since a history standard can be completely addressed through meeting the benchmarks at any one level. Thus, the number of standards in place at any one time is considerably less than the total. For a discussion, see Kendall & Marzano (1995).

[2]The national standards project for world history has identified those standards that are essential (core) and supplementary standards that may be used at discretion (related).

APPENDIX B

Bibliography of documents cited in the project's database

Bradley Commission on History in the Schools. (1988). *Building a history curriculum: Guidelines for teaching history in the schools.* Washington, DC: Educational Excellence Network.

California Department of Education. (1994). *Health framework for California public schools, kindergarten through grade twelve.* Sacramento, CA: Author.

California Department of Education. (1991). *Model curriculum standards: Grades nine through twelve.* Sacramento, CA: Author.

California Department of Education. (1990). *Science framework for California public schools, kindergarten through grade 12.* Sacramento, CA: Author.

California Department of Education. (1989). *Visual and performing arts framework for California public schools: Kindergarten through grade twelve.* Sacramento, CA: Author.

Carnevale, A.P., Gainer, L.J., & Meltzer, A.S. (1990). *Workplace basics: The essential skills employers want.* San Francisco, CA: Jossey-Bass.

Consortium of National Arts Education Associations. (1994). *National standards for arts education: What every young American should know and be able to do in the arts.* Reston, VA: Music Educators National Conference.

Crabtree, C., Nash, G. B., Gagnon, P., & Waugh, S. (Eds.). (1992). *Lessons from history: Essential understandings and historical perspectives students should acquire.* Los Angeles, CA: National Center for History in the Schools.

Geographic Education National Implementations Project. (1987). *K-6 geography: Themes, key ideas, and learning opportunities.* Washington, DC: Author.

Geography Education Standards Project. (1994). *Geography for life: National geography standards.* Washington, DC: National Geographic Research and Exploration.

Joint Committee on Geographic Education. (1984). *Guidelines for geographic education: Elementary and secondary schools.* Washington, DC: Association of American Geographers.

Joint Committee on Health Education Terminology. (1991). Report of the 1990 Joint Committee on Health Education Terminology. *Journal of Health Education,* 22, (2), 97-107.

Joint Health Education Standards Committee. (draft, 1994). *National health education standards and performance indicators.* Reston, VA: Association for the Advancement of Health Education.

Michigan State Board of Education. (1988). *Michigan essential goals and objectives for health education.* Lansing, MI: Author.

Music Educators National Conference. (1986). *The school music program: Description and standards.* Reston, VA: Authors.

National Assessment of Educational Progress. (1989). *Science objectives: 1990 assessment.* Princeton, NJ: Educational Testing Service.

National Assessment of Educational Progress. (1992). *Description of writing achievement levels-setting process and proposed achievement level definitions.* Iowa City, IA: American College Testing Program.

National Assessment of Educational Progress. (1992). *Mathematics assessment framework for* *the 1994 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board.

National Assessment of Educational Progress Arts Education Consensus Project. (1994). *Arts education assessment framework*. Washington, DC: National Assessment Governing Board.

National Assessment of Educational Progress in U.S. History. (1994). *Provisional item specifications*. Washington, DC: Council of Chief State School Officers.

National Assessment of Educational Progress Reading Consensus Project. (1990). *Assessment and exercise specifications: 1992 NAEP reading assessment*. Washington, DC: National Assessment Governing Board.

National Assessment of Educational Progress Science Consensus Project. (1993). *Science assessment and exercise specifications for the 1994 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board.

National Center for History in the Schools. (1994). *National standards for history for Grades K-4: Expanding children's world in time and space*. Los Angeles, CA: Author.

National Center for History in the Schools. (1994). *National standards for United States history: Exploring the American experience*. Los Angeles, CA: Author.

National Center for History in the Schools. (1994). *National standards for world history: Exploring paths to the present*. Los Angeles, CA: Author.

National Committee on Science Education Standards and Assessment. (draft, November 1994). *National science education standards*. Washington, DC: National Academy Press.

National Council for the Social Studies. (1994). *Expectations of excellence: Curriculum standards for social studies*. Washington, DC: Author.

National Council of Teachers of English. (1982). *Essentials of English: A document for reflection and dialogue*. Urbana, IL: Author.

National Council of Teachers of English. (1989). *Democracy through language*. Urbana, IL: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Pearsall, M. K. (Ed). (1993). *Scope, sequence, and coordination of secondary school science. Vol. 1. The content core: A guide for curriculum designers*. Washington, DC: National Science Teachers Association.

Project 2061, American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York, NY: Oxford University Press.

Project 2061, American Association for the Advancement of Science. (draft 1993, January). *Benchmarks for science literacy (Part 1: Achieving science literacy)*. Washington, DC: Author.

Standards Project for English Language Arts. (draft, February, 1994). *Incomplete work of the task forces of the standards project for English language arts*. Urbana, IL: National Council of Teachers of *English*.

The Secretary's Commission on Achieving Necessary Skills. (1991). *What work requires of schools: A SCANS report for America 2000*. Washington, DC: U.S. Department of Labor.

BIBLIOGRAPHY

American Federation of Teachers. (1985, September). Critical thinking: It's a basic. *American Teacher*, p. 21.

Carnevale, A. P., Gainer, L. J., & Meltzer, A. S. (1990). *Workplace basics: The essential skills employers want*. San Francisco: Jossey-Bass.

College Board. (1983). *Academic preparation for college: What students need to know and be able to do*. New York: College Entrance Examination Board.

Dewey, J. (1916). *Democracy and education*. New York: Macmillan.

Educational Policies Commission. (1961). *The central purpose of American education*. Washington, DC: National Education Association.

Futrell, M. H. (1987, December 9). A message long overdue. *Education Week*, p. 9.

Geographic Education National Implementations Project. (1987). *K-6 geography: Themes, key ideas, and learning opportunities*. Washington, DC: Author.

Geography Education Standards Project. (1993). *National geography standards*. Washington, DC: National Council for Geographic Education.

Geography Education Standards Project. (1994). *Geography for life: National geography standards*. Washington, DC: National Geographic Research and Exploration.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 93-104.

Hazen, R. M., & Trefil, J. (1991). *Science matters: Achieving scientific literacy*. New York, NY: Doubleday.

Joint Committee on Geographic Education. (1984). *Guidelines for geographic education: Elementary and secondary schools*. Washington, DC: Association of American Geographers.

Joint Committee on Health Education Terminology. (1991). Report of the 1990 Joint Committee on Health Education Terminology. *Journal of Health Education, 22,* (2), 97-107.

Kendall, J.S. & Marzano, R.J. (1995). *The Systematic Identification and Articulation of Content Standards and Benchmarks: Update, March 1995.* (Technical Report). Aurora, CO: Mid-Continent Regional Educational Laboratory.

Levine, D.V., & Associates. (1985). *Improving student achievement through mastery learning programs*. San Francisco: Jossey-Bass.

Mager, Robert F. (1962). *Preparing instructional objectives*. Palo Alto, CA: Fearon Publishers.

National Assessment of Educational Progress. (1992). *Item specifications: 1994 national assessment of educational progress in geography*. Washington, DC: National Assessment Governing Board.

National Committee on Science Education Standards and Assessment. (draft, November 1994). *National science education standards*. Washington, DC: National Academy Press.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

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, DC: Government Printing Office.

National Education Standards and Improvement Council. (1993). *Promises to keep: Creating high standards for American students. Report on the review of educational standards from the Goals 3 and 4 Technical Planning Group to the National Education Goals Panel.* Washington, DC: National Goals Panel.

National Science Board Commission on Precollege Education in Mathematics, Science and Technology. (1983). *Educating Americans for the 21st century.* Washington, DC: National Science Board Commission.

Project 2061, American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* New York: Oxford University Press.

Resnick, L. B. (1987). *Education and learning to think.* Washington, DC: National Academy Press.

Shanker, A. (1992, June 17). Coming to terms on world-class standards. *Education Week: Special Report.* p. S11.

Shavelson, R., Baxter, G., & Pine, J. (1992). Performance assessment: Political rhetoric and measurement reality. *Educational Researcher, 21,* (4), 22-27.

Shepard, L. (1993). *Setting performance standards for student achievement: A report of the National Academy of Education Panel on the evaluation of the NAEP trial state assessment: An evaluation of the 1992 achievement levels.* Stanford, CA: The National Academy of Education, Stanford University.

Spady, W. G. (1988). Organizing for results: The basis of authentic restructuring and reform. *Education Leadership, 46,* (2), 4-8.

The Secretary's Commission on Achieving Necessary Skills. (1991). *What work requires of schools: A SCANS report for America 2000.* Washington, DC: U.S. Department of Labor.

Tucker, M. (1992, June 17). A new social compact for mastery in education. *Education Week: Special Report.* p. S3.

Yoon, B., Burstein, L., & Gold, K. (n.d.). *Assessing the content validity of teacher's reports of content coverage and its relationship to student achievement.* (CSE Report No. 328). Los Angeles: Center for Research on Evaluation, Standards and Student Testing, University of California, Los Angeles.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership, 46,* (7), 41-47.

Wiggins, G. (1993, November). Assessment: Authenticity, context, and validity. *Phi Delta Kappan,* pp. 200-214.