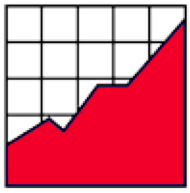


NATIONAL CENTER ON EDUCATIONAL OUTCOMES

CRAFTING CURRICULAR AIMS FOR

INSTRUCTIONALLY SUPPORTIVE ASSESSMENT

By

W. James Popham
University of California, Los Angeles

With Examples of Appropriate Curricular Aims

By

Roger Farr
Indiana University

and

Mary Lindquist
Columbus State University

February 2003

Support for the preparation of this document was supplied by the American Association of School Administrators, the National Association of Elementary School Principals, the National Association of Secondary School Principals, the National Education Association, and the National Middle School Association. No endorsement or approval of the contents of this document by those organizations should be inferred.

CRAFTING CURRICULAR AIMS FOR INSTRUCTIONALLY SUPPORTIVE ASSESSMENT

W. James Popham
University of California, Los Angeles

Not all curricula are created equal—at least for purposes of educational assessment. Some curricula can stimulate instructionally beneficial educational testing. Some can't. One of the most important challenges facing American educators today is how to conceptualize curriculum targets so that they are more likely to result in improved instruction.

This analysis describes how state curriculum can be optimally configured to foster *instructionally supportive assessment*; that is, assessment intended to promote more effective classroom instruction. The conception of instructionally supportive assessment offered here coincides with the views of the Commission on Instructionally Supportive Assessment as reflected in its October 2001 report[*] for educational policymakers. In that document, large-scale instructionally supportive tests were characterized as statewide assessments that would not only supply credible accountability evidence regarding the effectiveness of educators' instructional efforts, but would also support classroom instructional activities.

In an effort to assist state-level policymakers in adopting instructionally supportive tests, members of the Commission also supplied a document providing illustrative language that could be employed by states issuing a request for proposals (RFP) to test-development contractors.[**] The Commission hoped that by adopting or adapting the RFP-language supplied in that document, state assessment personnel could create an RFP that would lead to the construction of instructionally supportive assessments.

This analysis expands on the guidance the Commission provided in its model RFP by providing additional suggestions about how to sensibly prioritize a state's curricular aims. The suggestions are the author's alone, and should not be regarded as positions of the Commission on Instructionally Supportive Assessment.

In addition, the analysis contains two appendices that provide examples of the kind of curricular targets that would support the construction of an instructionally supportive accountability test. The first example, on purposeful reading, was developed by Roger Farr, Indiana University (see Appendix A); the second, on problem-focused number usage, was developed by Mary Lindquist, Columbus State University (see Appendix B).

---

[*] Commission on Instructionally Supportive Assessment. *Building Tests That Support Instruction and Accountability: A Guide for Policymakers.* Washington, DC: Author, 2001. Available online at www.aasa.org, www.naesp.org, www.principals.org, www.nea.org, www.nmsa.org.

[**] Commission on Instructionally Supportive Assessment. *Illustrative Language for an RFP to Build Tests That Support Instruction and Accountability*. Washington, DC: Author, 2001. Available online at www.aasa.org, www.naesp.org, www.principals.org, www.nea.org, www.nmsa.org.

An appropriate accountability test, of course, is one that can be used to secure accurate evidence about the instructional success of schools, districts, and states. A test that spurs improved classroom instruction is, in brief, a test that lets teachers:

1) know what skills and knowledge are to be assessed, hence the skills and knowledge they need to promote in their classrooms, and
2) determine, after instruction has been concluded, whether their teaching was successful.

Although the analysis focuses chiefly on the large-scale assessments now being used as the cornerstones of state-level accountability programs, it also is germane to the creation of instructionally beneficial tests for use in the classroom.

The Current Curricular Context

An Apprehension-Engendering New Law. American educators have been experiencing ever-increasing pressures during the last two decades to supply evidence regarding their effectiveness. That pressure was ratcheted up several notches on January 8, 2002 when President Bush signed into law the *No Child Left Behind* (NCLB) Act, which reauthorized the Elementary and Secondary Education Act.

This statute not only calls for a substantial expansion in the use of state-by-state academic assessments, it also requires educators to demonstrate *adequate yearly progress* (AYP) in the proportion of students displaying test-determined proficiency in their mastery of a state's curricular aims. Moreover, each state, district, and school must issue a progress-focused report each year—and the dominant factor in those report cards will be students' AYP as determined by each state's academic assessments.

The required levels of students' AYP called for in the NCLB Act are genuinely daunting. All schools failing to meet their state's AYP targets are to be publicly identified as deficient, and schools serving disadvantaged students (that is, schools receiving Title I funds) will soon be on the receiving end of a series of increasingly serious negative sanctions such as: (1) wholesale staff replacements or (2) being restructured as a charter or privately managed school. Accordingly, most clear-thinking educators are appropriately apprehensive about the potentially adverse impact of this far-reaching federal law.

In addition, the *No Child Left Behind Act* requires states to align their academic assessments with challenging academic content standards, that is, their approved curricular aims. These academic content standards are the skills and knowledge that authorized policymakers (typically, the state board of education) have approved as the instructional targets for educators and students. Unfortunately, in most states these standards are altogether unsuitable as a framework for the creation of instructionally supportive assessments.

Too many curricular targets. In most states, well-intentioned curricular specialists have generated a sprawling array of content standards which, far too often' list almost every imaginable sort of worthwhile (or, sometimes, marginally worthwhile) curricular outcomes.

Such "wish-list" content standards typically turn out to be far too numerous for assessment purposes—at least for real-world assessment during the time-periods actually allowed for such testing. In short, state content standards often consist of way too many curricular targets.

Even in states where there are only a modest number of content standards, those standards are often stated so generally that they must be explicated by "benchmarks" or "indicators" that, in turn, provide educators with some idea about curricular targets. This further explication, however, results in long lists of targets that overwhelm teachers and are too numerous to guide test-designers. Too many curricular aims underscore the considerable appeal of a "more-is-less" approach to crafting curriculum. But, too many curricular targets turn out to be no targets at all.

Unintended Consequences of Collegiality. Just to make sure that the curricular closet is completely free of skeletons, there is another reason (beyond the well-intentioned wish-listing described above) that state content standards are so numerous. More often than not, content standards are created by a group of subject-matter specialists and/or experienced teachers, typically the state's leading experts in various subject areas.

When these groups of subject-knowledgeable individuals get together, there is a built-in bias in favor of more—and more—content standards. Even if some members of the group regard a colleague's proposed content standard as less than wonderful, they are often reluctant to disagree. To oppose a colleague (and often a friend) is frequently difficult, if not impossible. It is typically easier for dissenters to silently "go along" and, thereby, add even more content standards to the ever-expanding pile of outcomes that teachers are supposed to promote.

Rarely does anyone say, "There are far too many things here for teachers to teach!" And yet, given a typical 180-day school year, there *are* almost always far too many content standards for teachers to teach. Indeed, in most states the huge numbers of state content standards oblige many teachers to address those too-numerous content standards with a superficial "lick-and-a-promise" approach to instruction. Unfortunately for students, the instructional "promises" are rarely kept.

Moreover, because assessment specialists are usually not invited to take part in these important curricular conclaves, seldom does one hear a comment such as: "There is simply no way we can measure all these content standards during the two-hour spring assessment period!" And yet, there is a widespread perception that statewide assessments can, in fact, be built to measure students' attainment of an enormous array of content standards.

Negative Consequences for Students. The net effect of all these events is that state-approved content standards, despite widely held views to the contrary, are almost always *too numerous to teach and too numerous to test*. Although typically created by well-intentioned educators who only want the best for a state's students, state content standards constitute a dysfunctional framework from which to build state-level tests. Given the nature of current content standards, those tests *cannot* be instructionally supportive. Yet, if instructionally supportive tests are *not* installed, then the sorts of instructionally *insensitive* tests now seen in most states' accountability programs will, without question, soon cause those educators to be regarded as instructional failures. This is because, as required by the NCLB Act, students will fail to make sufficient AYP

on state tests that are, by their very nature, staunchly resistant to detecting instructional progress, even if such progress were being made.

And, most importantly, some teachers—frustrated by being placed in the middle of a no-win accountability game—may adopt classroom practices that will seriously erode the caliber of education provided students. Students will be harmed, perhaps irreparably, by the presence of state-level accountability tests that are not designed to be instructionally supportive.

Such a negative conclusion can be avoided, however, *if instructionally supportive tests are installed* to function as states' academic assessments. According to the NCLB Act, these assessments must be aligned with states' content standards. Therefore, it is imperative to configure a state's content standards so that they can be appropriately assessed in a manner that benefits students instructionally. It can be done.

Key Attributes of Instructionally Supportive Accountability Tests

Instructionally supportive tests that satisfy the assessment and accountability requirements of the NCLB Act—and align with state curricula as configured in ways described in detail later in this report—have four key attributes:

They Assess Important Skills and/or Knowledge. An instructionally supportive accountability test focuses on a modest number of assessment targets—and the skills and/or knowledge being assessed must clearly be important. Good illustrations of significant yet measurable skills that have been suitably assessed over the years in most state accountability systems are students' composition skills.

The ability to generate an effective persuasive or expository essay is a powerful skill. And it can be measured appropriately in statewide tests. Indeed, at least in my view, America's students are far better writers today than they were a generation ago primarily because of high-stakes tests featuring the use of rubric-scored writing samples for assessing students' composition skills. Those high-stakes tests triggered the widespread adoption of highly successful classroom instructional strategies such as "the writing process" that emerged more than two decades ago from language arts specialists in the Bay Area Writing Project in San Francisco. In short, classroom instruction improved because of the installation of high-stakes tests measuring an important, *teachable* skill. In the writing-sample approach to assessment, teachers were able to address instructionally all the criteria in the scoring rubrics that were used to evaluate students compositions. The architects of the writing assessments kept teaching in mind.

But not all important assessment targets need to be measured via constructed-response tests, as is the case in writing. If, for instance, students need to learn a body of factual knowledge, and if their mastery is to be assessed by a state accountability test, then selected-response items (for instance, multiple-choice items) can usually do a decent job in determining the knowledge a student has mastered. An instructionally supportive test that assesses students' knowledge, however, must focus on an *important* body of knowledge.

Provide Assessment Descriptions. An instructionally supportive accountability test describes the skills and/or knowledge to be assessed so that teachers understand what is going to be measured. Moreover, such *assessment descriptions*—one description for each skill or body of knowledge to be assessed—must be succinct and easy to read so that busy teachers will be *willing* to read them. It is also almost always helpful if they contain an illustrative item or two. Examples often help teachers figure out what is being assessed.

The assessment descriptions that accompany an instructionally supportive accountability test must communicate clearly so that teachers understand precisely the curricular outcomes their students need to master. Teachers who possess a clear understanding of where they are to head instructionally will tend to arrive at that destination.

To determine whether an assessment description is sufficiently clear, all you need to do is ask a group of teachers to read the description, then *individually* (not in groups) describe in writing the nature of what is being assessed. Those individually authored descriptions ought to be similar and correct across teachers. If they aren't the assessment description is flawed.

Provide Standard-by-Standard Reporting of Test Results. Teachers who comprehend the nature of the curricular outcomes their students need to master (because of the availability of lucid assessment descriptions), will be in a position to provide effective instruction for their students. But was that instruction *actually* effective? If teachers are not provided with their students' test results on a standard-by-standard basis, then they are unable to discern which of their instructional activities are fine as-is and which ought to be altered. Standards-based tests without per-standard reporting constitute an educational charade. Moreover, any such per-standard reporting needs to be done at a student level, not just for classes, schools, or districts. In this way, additional instructional support can be given to those students whose mastery of particular content standards is unsatisfactory. Standard-by-standard reporting is a *sine qua non* for any kind of instructionally supportive accountability test.

Standard-by-standard reporting, however, usually requires more items or tasks per content standard than are typically included on most state accountability tests. Therefore, to provide standard-by-standard reporting, and avoid overly long tests, an instructionally supportive accountability test assesses fewer content standards than current state tests. Assessing a modest number of high-import curricular outcomes is far better for our students than attempting to assess a galaxy of content standards, and failing to report results for each of those standards.

Standard-by-standard reporting refers to the skills and knowledge that teachers will teach. In some states, a "standard" in a given subject is little more than a general label—and often a largely uninterpretable label. Several states, for instance, use the term "standard" to refer to such student outcomes as "listening" or "reading." In other states, "standards" refer to broad content categories such as "geometry" or "physical geography." Clearly, such "standards" are far too general to be serviceable for standard-by-standard reporting—that is, reporting students' test performance at levels of generality suitable for teachers' instructional decision-making.

While it may be instructionally hypocritical to claim that a seemingly endless galaxy of content standards can be properly taught in 180 days, it is equally hypocritical to contend that the same

array of content standards can be properly assessed in an hour or two. Honest accountability assessment is predicated on an instructional essential—providing of standard-by-standard results. This, in turn, requires measuring fewer curricular aims—more likely in the neighborhood of a half-dozen content standards than the 50 or 100 curricular aims allegedly measured in many states.

Assess Teachable Skills and Knowledge. Finally, the skills and/or knowledge to be assessed by an instructionally supportive accountability test need to be *teachable*. Teachers must be able to promote students' mastery of what is being measured. This is not to suggest that it should be fool's play for any teacher to whip up effective lessons aimed at the content standards that are assessed by instructionally supportive tests. Remember, the curricular targets being measured must be truly important ones and, as such, they will often be tough to teach. Accordingly, in many instances, teachers will need to be provided with considerable professional development to expand their repertoire of instructional tactics. Given such support, however, all teachers should be able to promote students' mastery of the skills and knowledge that are assessed by instructionally supportive accountability tests.

Indeed, the *potential teachability* of what is to be assessed should be kept constantly in mind by test developers. They should continually ask themselves, "How likely is it that teachers can get their students to master this content standard that we're trying to assess?"

To review, the four attributes of an instructionally supportive accountability test are the following:

- *It assesses truly important curricular aims.*
- *Its content is clearly set forth in teacher-palatable assessment descriptions.*
- *Its results are reported, standard-by-standard, at the student level.*
- *What it measures can be effectively taught by teachers.*

Preliminary Considerations Related to Configuring Curricular Aims

It is possible to conceptualize a state's curriculum in an *assessment-friendly* fashion so that instructionally supportive accountability tests can be aligned with that curriculum. But, prior to doing so, those who are involved in configuring curricular aims in an instructionally beneficial manner, need to consider a few preliminary issues.

Already-Approved Content Standards. The advice this paper offers about configuring curricular outcomes is not offered to educators existing in a vacuum. Most states already have officially adopted content standards, sometimes approved by a state board of education only after arduous struggles. Therefore, states should *not* try to replace their existing curricula or even modify them.

In fact, these existing curricula may provide some useful instructional guidance to a state's educators. (Although, in truth, I have serious doubts about the actual instructional payoffs of

most of today's wish-list curricula.) If an existing state curriculum seems to be educationally helpful, leave it alone. Use any such curricula as springboards for the generation of aligned and instructionally helpful assessment instruments.

Putting it another way, states should retain their currently approved curricular structure, but develop a supplementary *assessment framework* that, functioning as a companion document to already approved content standards, makes clear how the state's accountability tests will be aligned with those content standards. Brief alignment justifications could be supplied in the assessment framework. They would show how any to-be-assessed skills and bodies of knowledge are derived from, and therefore aligned with, certain of the state's existing content standards. Given the generality of most existing state content standards, such justifications can almost always be provided.

To put this in policy terms, for the NCLB Act to have a positive rather than negative effect on education, states need instructionally supportive accountability tests. To create such tests—and tests that are demonstrably aligned to already approved content standards—it is *not* necessary to undertake a Herculean re-working of state curricula. Rather, states can develop an accompanying, instructionally oriented assessment framework that provides teachers with assessment-grounded instructional targets.

Time Available for Assessment. Those who are involved in configuring curricular aims in an instructionally beneficial manner need to ask, "How much time is actually available for the administration of our state's accountability tests?" Clearly, if a 50-minute period is available for assessing students' mastery in a given subject—at a given grade level—then the assessment should focus on fewer content standards than could be accommodated if two or three such assessment periods were to be available. But opting for *more* assessment is not necessarily the direction in which state policymakers ought to go. Clearly, the more time that students spend being tested, the less time will be available for them to be taught.

Because the NCLB Act calls for annual assessments in reading (or language arts) and mathematics in grades 3-8, it is apparent that—at minimum—these subject areas must be assessed each year. Later, the law calls for assessments in science in certain grade ranges. This is a substantial amount of testing that will require state department of education personnel responsible for curriculum *and* assessment to exercise sound, sophisticated judgment in carrying out *collaboratively* what will surely be a nontrivial balancing act. They will have to choose between (1) increased testing-time that assesses more of the state's high-import content standards, but eats up important instructional time, and (2) reduced testing-time that intrudes less on classroom instructional time, but fails to assess students' mastery of some genuinely important curricular outcomes.

Constructed-Response Assessment. Another important, practical issue to be addressed early deals with the amount of constructed-response items that can be included in an instructionally supportive accountability test. Clearly, students who can generate "from scratch" responses to challenging tasks will most likely have mastered a given content standard more completely than will students who can only select correct answers from sets of already-presented options.

Yet, the scoring of students' constructed-responses is costly and must be done carefully in order to maintain satisfactory levels of scorer-accuracy. Some states (e.g., Kentucky and Maryland), only a few years ago, embarked with considerable acclaim toward creating state assessment programs containing substantial numbers of constructed-response items. Recently, for both fiscal and psychometric reasons, many states have now retreated to the use of tests dominated by selected-response items.

Curricular and assessment specialists need to know if they have the freedom to conceptualize skills in a fashion that will require constructed-response measurement. It would be ludicrous to allow a curriculum to be configured for largely constructed-response assessment, then learn that such a measurement approach was precluded because of budgetary constraints. The people who take on the task of creating an assessment framework that is aligned with state curricula need to know any important practical limits that govern their deliberations.

Grade-To-Grade Articulation. When a number of adjacent grades are to be tested, as in the case with the six adjacent grades (3-8) required by the NCLB Act, it is a perfect opportunity to make the grade-to-grade articulation of content standards as instructionally sensible as possible. Necessary earlier skills can be introduced in an educationally efficient fashion. Curricular redundancies can be avoided. Thus, any curriculum and assessment personnel wrestling with these important problems must include persons capable of isolating a coherent series of incremental curricular steps between grades 3 and 8 in the subjects that will be assessed through the use of instructionally supportive accountability tests.

Current and Future Funding. Several fiscal questions, as prosaic as they may seem, need to be considered well in advance of configuring a state's curriculum for an assessment framework. How much money is available for test development? How much money is available for test administration, scoring, and reporting? Is there money for professional development for teachers that is based directly on instructionally supportive accountability tests? How much federal money will be available and how much state money will be available? How long is any such funding apt to exist? Is it possible for several states to pool their federally supplied test-development dollars to create test-development consortia?

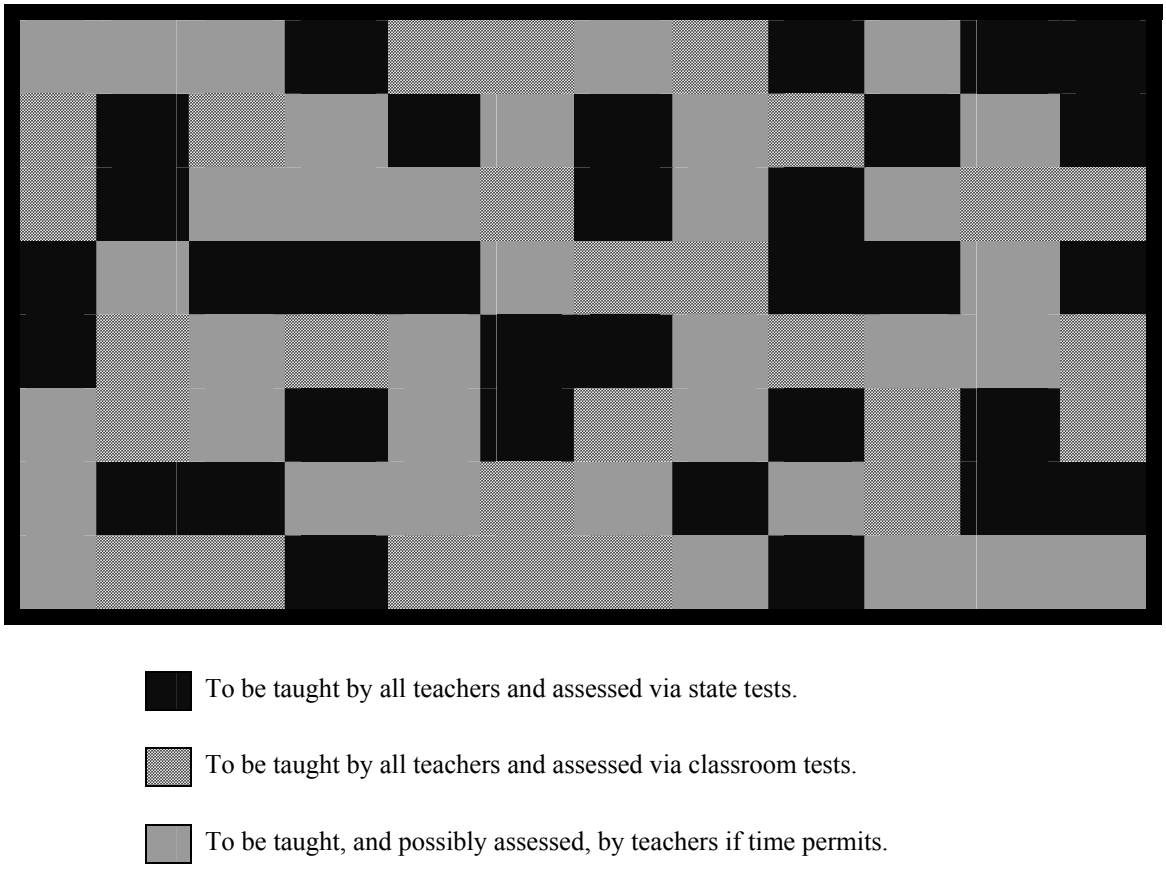
Configuring Curricula for Instructionally Supportive Assessment

Configuring a curriculum's skills and knowledge so they serve as a suitable springboard for creating instructionally supportive accountability tests centers on a key task—capturing as much as possible of the most important things for students to learn, yet doing so in a manner that conforms to instructionally supportive assessment. This is a substantial undertaking.

Re-Conceptualizing the Curriculum. As a first step, states need to re-conceptualize many small-scope curricular aims into larger-scope curricular aims. In other words, the challenge is re-configuring multitudinous content standards into a smaller number of broader, significant, *yet teachable and assessable* content standards.

With the hope that graphic representations of this process may be of some help, consider Figure 1 in which the large, heavily bordered rectangle is intended to represent a state's approved curriculum in a given subject (for instance, mathematics) at a particular grade level or grade range.

Figure 1. A TYPICAL STATE-APPROVED CURRICULUM



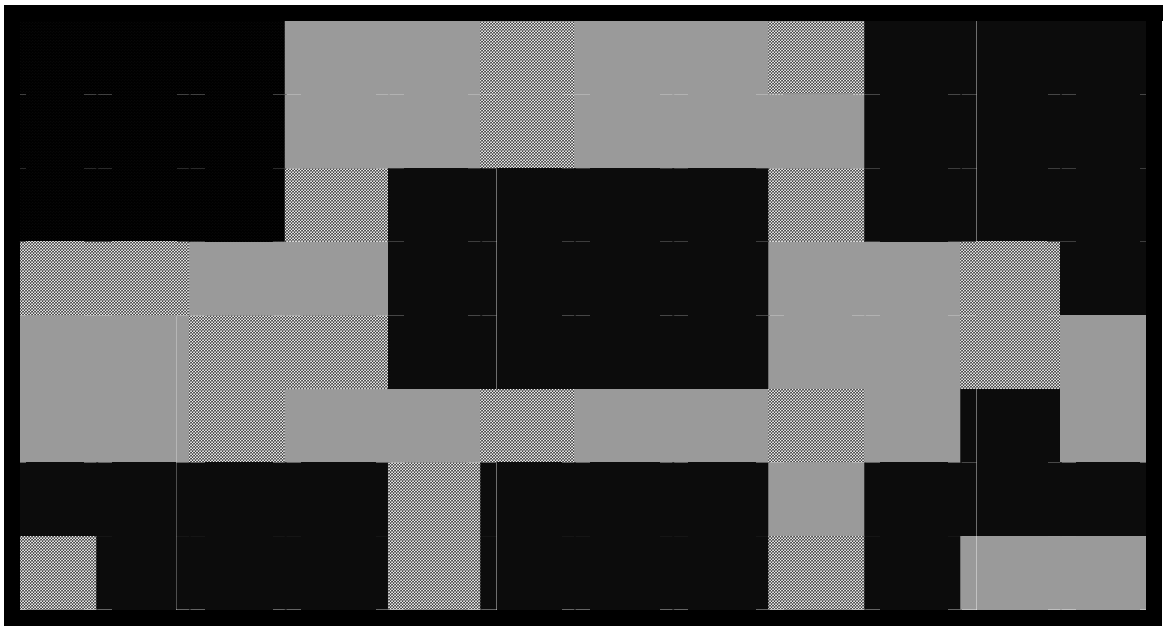
Each of the small geometric shapes within the rectangular curriculum's borders represents a state-approved content standard. Three categories of content standards are depicted. The *black* content standards are those considered sufficiently important to be taught by all teachers and assessed annually on a statewide test. The *dark-gray* content standards, somewhat less important, are to be taught by all teachers, but measured via teachers' classroom assessments rather than a statewide test. The *light-gray* content standards, less important though still desirable, are to be taught (and possibly assessed) by teachers if time permits during the year.

In some states any such instructional and assessment expectations associated with each of the state's content standards have been explicitly made and widely promulgated to educators. In

other states, these assessment/instructional designations have either not been made or, if made, have not been disseminated.

Figure 1 contains 54 separate content standards (19 black, 16 dark-gray, 19 light-gray)—too many to be assessed during any reasonable test-taking time so that per-standard results can be reported. Furthermore, there are far too many content standards for most teachers to attend to during their day-to-day instructional activities. The task, then, is to re-configure those too-numerous, typically small-scope content standards into a lesser number of more suitable (for assessment) curricular targets such as can be seen in Figure 2.

Figure 2. AN ASSESSMENT-FRIENDLY STATE CURRICULUM



To be taught by all teachers and assessed via state tests.

To be taught by all teachers and assessed via classroom tests.

To be taught, and possibly assessed, by teachers if time permits.

In Figure 2 there are now only six broad-scope content standards to be assessed on a state test. There are only 12 content standards (dark gray) to be assessed via teachers' classroom tests and 10 content standards (light gray) to be assessed by teachers "if time permits." The content standards to be measured on the state test (black) have been deliberately re-conceptualized for statewide assessment purposes. Indeed, even those content standards that are slated for classroom assessment have been re-conceptualized so they are broader and few in number.

Focusing on the half-dozen or so broad-scope content standards that have deliberately been reconfigured for state-level assessment purposes, it is then possible to create an instructionally supportive test that can be an accurate accountability assessment, yet also function as a boon to teachers (and, therefore, to their students).

A Curricular Reconfiguration Group. As a next step, states need to convene curriculum and assessment specialists who function as a Curriculum Reconfiguration Group (CRG)

The CRG should include the state's most insightful and creative curricular specialists, educators who understand their subject areas inside-out, and can make sound judgments about the conceptual integrity of any reconfigured content standards. But, of equal importance, these educators must also be able to predict whether teachers can address any newly configured curricular aims in their instruction. The CRG also should include assessment-knowledgeable individuals who can provide guidance regarding the technical requirements of large-scale testing.

The CRG is charged with the task of devising an assessment framework that will serve as a template for the creation of instructionally supportive statewide accountability tests. The framework is derived from and aligned with the state's existing content standards. To accomplish this goal, the CRG must try to isolate important skills or bodies of knowledge akin to those represented by the six large (black) geometric shapes in Figure 2.

Ideally, the CRG should seek skills or bodies of knowledge that (at least with respect to their size) are cognitively comparable to students' composition skills (as assessed by writing samples). If the content standard deals exclusively with the student's mastery of factual knowledge, then that body of knowledge should be sufficiently substantial and significant to warrant its assessment on a high-stakes accountability test. More often than not, successful movement toward suitably large content standards requires a *coalescence* of smaller-scope content standards into larger ones. This problem revolves around what has usually been referred to as "grain size," that is, the breadth of curricular targets.

The process by which a given subject area's curricular aims can most efficiently be reconfigured typically depends both on the subject area itself and on the particular individuals (curricular and assessment specialists) who are members of a given CRG. A substantial amount of starting, stopping, and revising will be required. This is because this task often requires participants to abandon traditional grain-size curricular conceptions and move, instead, toward more innovative, encompassing ways of thinking about curricular targets. And all the while, of course, *instructional implications* need to govern the CRG's deliberations.

Four Evaluative Criteria. As the CRG decides on suitable content standards for a state-level assessment framework, it can apply the following four evaluative criteria to reconfigured standards. Not surprisingly, these criteria are closely allied to the four attributes of an instructionally supportive accountability test described earlier.

Typically, the application of the four evaluative criteria will oblige the CRG to jettison some contending content standards and to seriously re-work others. The enterprise is apt to be highly iterative and replete with a host of false starts and en route modifications.

However, if the resulting curricular aims have *not* been fashioned in a form suitable for instructionally supportive assessment, then it is certain that instructionally supportive accountability tests will not emerge. And this would be unfortunate because students will be the real losers in any state whose accountability tests are instructionally unsupportive. Our students deserve instructionally supportive tests.

- **Evaluative Criterion 1: Importance.** *A content standard must represent an unarguably significant skill and/or body of knowledge*.

Because an instructionally supportive accountability test will assess a relatively small number of content standards, every to-be-assessed curricular target must deal with a really significant educational outcome. Thus, the first evaluative criterion focuses on the educational importance of any reconfigured curricular aim.

Members of the CRG can provide anonymous independent ratings regarding the educational importance of each reconfigured content standard under consideration. Such ratings should be tabulated and made available during the content-reconfiguration process to all CRG members so that en route modifications of ratings can be made. Clearly, those content standards receiving negative or equivocal importance evaluations would need to be modified or, possibly, abandoned.

- **Evaluative Criterion 2: Describability.** *A content standard must be capable of being described so that teachers (1) can understand what is to be assessed and (2) will be apt to read the content standard's assessment description*.

The CRG must create written descriptions for teachers of the skill and/or knowledge assessed in each content standard. This assessment description should be sufficiently succinct and clear that busy teachers will be willing to read it. (Incidentally, if a content standard deals only with a student's mastery of factual knowledge, then the *complete* set of knowledge eligible for testing needs to be made available—typically in the form of a separate supplement.) One or more illustrative items should be included, to make the assessment description more concrete and useful to teachers.

To make certain that any assessment description is apt to be read by teachers, teachers should be asked to review it. (Teacher members of the CRG would not be suitable for such tryouts because of their familiarity with the content standard.) Typically, a small group of teachers would be given a draft of an assessment description, then *individually* asked to set forth—orally or in writing—what they understand the nature of the content standard to be. If those individual interpretations of the content standard fail to be accurate and homogeneous, then the assessment description will need to be revised.

Those same teachers should also be asked to supply completely *anonymous* indications regarding the likelihood that a busy teacher will actually read the assessment description. If many teachers indicate their colleagues will be unwilling to read the draft assessment description, then it must

be modified. Even lucid assessment descriptions, if unread by teachers, will not benefit instruction.

- **Evaluation Criterion 3: Per-Standard Reportability.** *A content standard must be measurable so that reasonably accurate standard-by-standard reporting can be provided for individual students.*

To supply reasonably accurate estimates of a student's content-standard mastery, it will usually be necessary to employ more items or tasks per content standard than is typically seen in today's achievement tests. On the other hand, assessment programs now used in many states employ only a *single* writing sample to arrive at decisions regarding a student's composition skills. So, clearly, the number of items or tasks needed for a to-be-assessed content standard depends directly on the particular content standard to be measured. Members of a CRG will need to give this issue careful consideration, then come up with their best judgment about the number of items needed per content standard.

At this point in the reconfiguration process, the CRG's assessment specialists can be very useful in determining how many items are needed per content standard. The quest in this instance, however, is for *reasonably accurate* inferences about a student's standards-mastery—*not perfectly accurate* inferences. Psychometrically trained specialists characteristically yearn for unflawed measurement tools. But such unflawed measurement tools typically turn out to be onerous because of their excessive length. This is an instance where praiseworthy assessment aspirations must be trumped by practical education realities.

As with most important issues in education, trade-offs must be made between measurement precision and instructional utility. My own preference in these conflicts is to favor instruction over assessment, but members of each CRG will need to reach their own conclusions. Remember, the task is to create *instructionally supportive tests.* If the instructional yield of these tests evaporates because of a quest for measurement precision, then the tests will not benefit teachers or students.

The essence of this third evaluative criterion is quite straightforward. An emerging content standard must be capable of being reported so that an *individual* student's mastery of that standard can be determined reasonably well. Teachers, students, and students' parents need to find out which content standards a student has mastered and which the student has not mastered. If a curriculum's content standards cannot be reported for individual students in a per-standard manner, then the curriculum needs more reconfiguring.

- **Evaluative Criterion 4: Teachability.** *A content standard must be instructionally addressable by ordinary teachers so that students' mastery of that curricular aim can be promoted.*

The reconfigured content standard must be teachable—and teachable by regular teachers, not siblings of Socrates. Ideally, teachers will be given ample professional development dealing with alternative instructional tactics they might wish to employ in promoting students' mastery of each content standard included in the assessment framework. However, the CRG needs to

wrestle seriously with the likelihood that run-of-the-classroom teachers can get their students to master a given content standard. If that kind of instructional success is unlikely, then the content standard needs to be reconfigured once more.

The way that this final evaluative criterion is addressed by a CRG will typically be something like this: The group's instructional specialists must be constantly thinking, "Could I teach my students to master this skill (or body of knowledge) as it is currently conceptualized?" An even more potent way for a CRG member to think about a content standard is to ask this question, "If required to do so, could I help a group of typical teachers learn how to successfully promote their students' mastery of this content standard?" If negative answers are given to either of these questions, then more reconfiguring is required.

Examples of Appropriate Curricular Targets

Two exemplars of curricular targets on which test developers can build instructionally supportive assessments, and useful assessment descriptions for teachers, are included in the appendices. In Appendix A, Roger Farr of Indiana University provides an example of a curricular aim in reading that would be suitable for the construction of an instructionally supportive accountability test. In Appendix B, Mary Lindquist of Columbus State University supplies an illustration of a curricular aim that would be suitable for the creation of an instructionally supportive test in mathematics.

***Purposeful Reading*.** The reading example in Appendix A focuses on a potent reading skill described as *purposeful reading*. The curricular aim rests squarely on the premise that not only do people read for a purpose, but that they *should* read for a purpose. Moreover, for different kinds of text materials, a reader's purposes will vary. If students can be taught to consciously determine their reading purposes in advance, then evaluate the success with which they have accomplished those purposes on the basis of purpose-specific evaluative criteria, they will become better readers. And, happily, as Farr has crafted this curricular outcome, the purpose-specific evaluative criteria to be employed in judging a student's success are all derivative from three more general, unchanging evaluative criteria.

This *purposeful reading* curricular target can be conceptualized into an eminently teachable framework. In my view, if *purposeful reading* were assessed and (as a near-certain consequence) actively taught in a state, the pursuit of this reading skill could have the same powerful impact on children's reading comprehension that the introduction of writing-samples accompanied by instructionally relevant rubrics has had on children's composition skills. (Incidentally, because of his considerable conversance with the way that standardized achievement tests are constructed, Farr also addresses assessment issues in Appendix A.)

Figure 3 provides an overview of *purposeful reading*. There are distinctive purposes for reading three types of text: *functional* (such as a train-schedule), *expository* (such as a section of an encyclopedia), and *narrative* (such as a mystery novel). Students' responses to assessment items would be evaluated on the basis of three general evaluative criteria: *accuracy, relevance,* and

sufficiency. These three criteria are particularized to some extent according to the particular text-linked purpose involved.

Figure 3. A Framework for Evaluating Purposeful Reading

PURPOSEFUL READING

Purposes for Reading *Functional Texts*

1. Determine information's relevance and importance.
2. Select and apply information for a task.

Purposes for Reading *Expository Texts*

1. Understand main points and supporting details.
2. Recognize expositional organization and its use.
3. See relationship of text's content to broader issues/topics.

Purposes for Reading *Narrative Texts*

1. Identify the development of character, setting, and mood.
2. Understand a story's plot development.
3. Identify a story's theme(s) and its (their) development.

Student's responses are scored on the basis of text-by-purpose specific applications of three general evaluative criteria: *accuracy, relevance,* and *sufficiency.*

A state test, then, might include items in which the student was *initially* given a purpose for reading a text selection. For example, the item might say something such as, "Suppose you were reading the following information from the U.S. Internal Revenue Service in order to determine if you needed to pay taxes this year." Then, having identified a purpose for reading, a segment of functional text would be presented in the exam—followed by questions (selected-response *or* constructed-response) that would elicit responses to permit inferences about whether the *test-given* purpose for reading had been accomplished.

Incidentally, care would need to be exercised in devising items related to this skill so that an item's mathematical demands were not excessive. For instance, in Roger Farr's examples of items associated with functional texts, the item-writers would have to make sure that the quantitative loadings of such items were not so great that the mathematical content reduced the validity of inferences about a student's mastery of the purposeful reading skill.

Classroom instruction aimed at promoting the student's mastery of this *purposeful reading* skill would require students to initially isolate the specific purpose for which they were reading a given type of text, then use text-by-purpose evaluative criteria to determine if their purpose had been accomplished. Students who master this powerful skill will be good readers!

Problem-Focused Number Usage. The mathematics example in Appendix B is described as *problem-focused number usage.* This curricular aim addresses a child's number usage mastery, but in relation to solving problems. Indeed, it is because of the considerable importance of problem-solving in mathematics that this curriculum outcome deliberately places students' mastery of number-usage *skills* and *concepts* in problem solving contexts. The number-usage curricular aim deals with *whole numbers*, *fractions*, and *decimals*.

Conceiving of *problem-focused number usage* as a content standard, the material in the illustrative curricular aim focuses directly on explanations and content delimitations suitable for classroom teachers. The curricular target provides a general assessment description suitable for all grade levels, then offers grade-specific assessment descriptions for two grades, each of which are accompanied by illustrative items that could be used to measure students' mastery of this powerful content standard.

Figure 4 illustrates the approach used to delineate this powerful mathematics skill, focusing in particular on the descriptive strategy used for the *skills* and *concepts* to be assessed for grade four. As the Figure shows, this approach to creating assessment descriptions in a *"content-rich"* field, such as mathematics, relies on the isolation of particular skills and concepts associated with three kinds of numbers deemed appropriate for fourth-grade students.

Figure 4. A Framework for Evaluating Problem-Focused Number Usage

PROBLEM-FOCUSED NUMBER USAGE

Content Standard: *Students can use numerical skills and concepts needed to solve routine and non-routine grade-appropriate problems.*

Skill Mastery will be assessed by determining if a student can (1) perform the skill in isolation or (2) use the skill in a problem-solving situation.

Concept Mastery will be assessed by determining if a student can (1) use the concept appropriately or (2) discriminate between proper/improper concept usage.

Whole Numbers:

Concepts: (Identified at each grade level)

Skills: (Identified at each grade level)

Fractions:

Concepts: (Identified at each grade level)

Skills: (Identified at each grade level)

Decimals:

Concepts: e.g., for grade 4: (1) place values—two places, (2) relation of decimals to whole numbers, fractions, and money

Skills: e.g., for grade 4: (1) read, write, and order decimals—two places, (2) add and subtract decimals—two places

Clearly, as a state's mathematics educators isolated the skills and concepts that were critical at different grade levels, they would employ the same sort of less-is-more curricular thinking that should be used when wrestling with myriad content standards. For every additional skill or concept that is added at a given grade level, there will be a corollary reduction in teachers'

instructional attention. A modest number of preeminently important skills and concepts can be, and if measured on a high-stakes state test, will be taught. A huge number of skills and concepts will overwhelm—everyone.

The illustrative items in Appendix B employ both *task-specific* scoring rubrics as well as *general* scoring rubrics. Although such rubrics clearly do not possess the instructional power of the sorts of *skill-focused* rubrics used when scoring students' writing samples, this approach is a more reasonable one to employ in fields such as mathematics.

***Wrap Up*.** Soon most state officials will be deciding how they will proceed in satisfying the new federal assessment requirements. The easiest way to do so, of course, is to buy the same old traditionally constructed achievement tests. But, though easiest, such a choice is neither in the best interest of educators nor, more importantly, of students.

Instructionally supportive tests can supply accurate accountability evidence—and also improve the quality of education that students receive. But instructionally supportive tests must be based on a suitable collection of curricular aims. The foregoing analysis suggests how a state's content standards can be made more assessment friendly.

Appendix A
Purposeful Reading

Roger Farr
Chancellor's Professor Emeritus
Indiana University

Reading is a purposeful activity. Teachers help students learn to read as a process of gaining meaning from text—and then applying what has been comprehended to complete an activity of some sort. This instructional focus is not surprising because all readers read for a purpose both in and outside of school.

There are many purposes for reading: enjoying the intrigue of a mystery, understanding how to solve a problem, reading the directions for setting up a new computer, gathering information needed to perform some repair, taking part vicariously in an adventure, learning about interesting characters, or gaining new knowledge about science or history or the contribution of someone who made a difference in one of those fields.

Good reading instruction focuses the attention of readers by teaching them what to expect from different kinds of texts, how those texts can fulfill their needs (their purposes for reading), *and* the strategies for constructing the meaning they need from particular kinds of texts. Students learn that different purposes require different kinds of reading and involve different types of texts. To know if they are reaching this goal, teachers need assessments designed to reveal their students' abilities to comprehend the types of texts related to typical and reasonable purposes that readers may have.

Different Types of Texts

If valid reading purposes are served by different types of texts, both instruction and the assessment of its effectiveness, must be based on the kinds of texts that readers read to accomplish purposes. There are three basic types of texts that students will encounter both in school and in everyday life:

Functional Texts: These texts include reading materials such as directions, schedules, maps, diagrams, and explanations for doing something or getting somewhere. They provide basic information that readers need to accomplish day-to-day tasks. Overriding strategies for making the most effective use of such texts are to skim, looking for information that serves a specific need, *or* to read carefully, considering and evaluating the usefulness of all details as in following directions.

Expository Texts: These texts include such things as textbooks, encyclopedias, biographies, scientific explanations, and historical and political analyses. These are usually read to learn new information that increases a reader's understanding of some topic.

Narrative Texts: These texts include stories, poems, novels, plays, and essays that are read to learn about people, to vicariously experience the characters and settings, to escape to imaginary places and times, and to become absorbed in adventure and fictional events, and various problems and solutions that structure the plots of these texts.

Purposes for Reading

If reading is done to accomplish specific purposes, then assessments are needed to help teachers plan instruction, and the assessments need to cover a reasonable range of those purposes. Because functional, expository, and narrative types of text are read for different purposes, assessments will ideally cover a reasonable number of the reader's purposes and, therefore, they will necessarily include a variety of text genres.

The following list of purposes for reading is not meant to be comprehensive. Indeed, it would probably be impossible to develop such a list. This list, however, can guide the development of assessments and, consequently, classroom instruction:

Functional Texts

Purpose 1: Read to determine the relevance and importance of functional information.

Purpose 2: Read to select and apply relevant information for a given task.

Expository Texts

Purpose 1: Read to understand a text's major points and supporting details.

Purpose 2: Read to understand the text's organizational exposition and how that organization serves the writer's purpose.

Purpose 3: Read to understand how the information in the text fits into broader topics and issues.

Narrative Texts

Purpose 1: Read to recognize and understand an author's development of character, setting, and mood as basic story elements.

Purpose 2: Read to understand how the plot of a story develops as a series of high points and/or how it can be depicted as a problem and its solution.

Purpose 3 Read to understand the theme of a story and how the author develops it.

Criteria for Evaluating Student Responses

If the assessments designed to measure purposeful reading comprehension across types of texts are to effectively guide instruction, they need to clearly identify the evaluative criteria to be used in judging successful completion of the assessment activity. If they do not do this, teachers are left with meaningless numbers that cannot be used to guide effective reading instruction.

The assessment plan needs to provide a careful description of what is meant by success. That description, of course, has obvious implications regarding the kind of classroom instruction that will improve reading as a purposeful activity. An assessment based on reading purposes would require distinctive criteria for the accomplishment of each purpose. These specific evaluative criteria are based on three general evaluative criteria. The more general evaluative criteria provide a guide to the development of more specific evaluative criteria for each purpose. These general evaluative criteria are:

- **Accuracy**: How *accurate* is the reader's grasp and use of the text?
- **Relevance**: How *relevant* is the textual detail or understanding the reader uses to fulfill the purpose?
- **Sufficiency**: Does the reader demonstrate and use a *sufficient* amount of the text to fulfill the task?

For example, if a reader has read several expository texts about global warming and is asked to report what the texts describe as "global warming," a successful accomplishment of the task would include *accurate* reporting of what the scientists quoted say about the concept; the student response would select information from the texts that is most *relevant* to the task of preparing the report, and it would use enough of the information from the texts to explain the concept—it would rely on the texts to a *sufficient* extent.

Thus, evaluation of student responses to indicate the success of their use of needed reading strategies could be evaluated using these evaluative criteria:

Does the reader *accurately* report the information in the texts?

Does the reader select information that is *relevant* to complete the purpose/task?

Is the *extent* of what the reader provides *sufficient*?

Putting It All Together

If instructionally supportive assessments are to be developed following the outline described above, those who construct them must be careful to create reading tasks that validly

reflect the general reader purposes presented. The assessment must include realistic reading activities for students, and they must provide the teacher with needed information that can be the foundation of instructional activities in a typical classroom.

What would such assessments look like? Each assessment activity would include the following minimal components:

1. A specific purpose for reading based on the type of text being read.
2. One or more reading text(s) that resembles realistic reading activities.
3. Assessment activities that respond to the purpose posed to the student at the beginning of the selection. These activities could be either student constructed responses (open-ended items), student selected responses (multiple-choice items), or a mix of the two.
4. Evaluative criteria for judging the reader's responses to the assessment activities.
5. Optional instructional suggestions that can help students perform better on the kinds of reading activities represented by the assessment.

Three examples of the specifications for such passages are provided below to show how such assessments could be developed. One for each general text type is presented, using one of the purposes above for that type of text.

Specifications for an Assessment Using a FUNCTIONAL Text

Purpose: ***Read to select and apply relevant information for a given task.***

1. Specific reading purpose: The reader would be asked to use a particular basketball team's schedule and the basketball league standings to determine whether the basketball team has a chance of winning the conference basketball championship.

2. Reading text: The test would include the following two items:

- A basketball schedule for a school team. The schedule would indicate whether the games were *at home* or *away* and would give the time the games would be played and the dates. It would include scores for games already played.
- The standings of all of the teams in the school's conference. The table showing the conference standings would list the teams in order of their records to date and give their win/loss records. It would also indicate how many games each team has yet to play.

3. Assessment activities:

- A student-constructed response (open-ended item) would ask the student to tell how good the school's chances are for leading the conference at the end of the season and to provide the evidence from the schedule and the conference standings to support the conclusion.
- A student-constructed response would ask the student which other schools in the conference stand a chance of winning the conference championship and why the reader

thinks that team could win the conference.

- A student-selected response (multiple-choice item) would ask which remaining game on the schedule is apt to be the most important in determining the conference champion.
- Other student-selected responses could (1) pose questions about details, such as dates of games, particular scores, particular school records; (2) ask which game played was the closest contest; (3) require the reporting of relevant details in the conference standings.

4. Criteria for judging/evaluating responses: The student-constructed responses would be judged on the basis of *accuracy, relevance*, and *sufficiency*. The evaluation would be based on whether the response:

- demonstrated that the reader had used information from the schedule and team standings accurately;
- demonstrated that the reader had selected information that was relevant to the task and that it was applied logically and validly; and
- whether sufficient information and details from the text had been selected to complete the task adequately and appropriately.

The student-selected responses would be written with distractors based on these same criteria.

5. Instructional suggestions: The scoring of the responses should lead to an indication of how the teacher can reinforce reading strategies that were relevant to the activity. The kinds of instructional activities that would follow from the results of this assessment might emphasize:

- Reading schedules, reports, and similar texts that provide information that must be synthesized to be understood.
- Identifying details in texts supporting conclusions that were stated or could be drawn from them.
- Skimming functional reading selections to synthesize or compare information across two or more texts.

Specifications for an Assessment Using an EXPOSITORY Text

Purpose: ***Read to understand a text's major points and supporting details.***

1. Specific reading purpose: The student would be asked to read two articles about skateboarding and to identify the reasons that various groups think a skateboard park should be constructed with city funding.

2. Reading text: The text consists of two passages:

- The first is a report from a city council meeting where many citizens complained that the city should not spend money to build a skateboard park. They argued that the costs are too high and the city cannot afford another park. Some citizens complained that the park

will be a place where students just gather and get in trouble. Some city officials reported on the dangers of skateboard accidents. A city attorney explained that insurance premiums would be very high to cover the city's liabilities if a skateboard park is built.

- A second article is a request from a group of citizens for the establishment of a city-run skateboard park. The article argues that children learn important physical skills and self-confidence from mastering skateboarding. They describe how students need a place to skateboard so they are not causing problems on streets and sidewalks. The article also discusses the need for more recreational activities in the city and how popular skateboarding is with teenagers. The article provides cost figures for developing a skateboard park that has been built and successfully maintained in another city.

3. Assessment activities:

- A student-constructed item would ask the students to identify the major criticisms of skateboarding and to provide details that support those conclusions.
- A student-constructed item would ask the student to compare the major arguments provided against the construction of a skateboarding park and the major reasons that people think such a park is a good idea.
- Student-selected items would focus on details that support the risks and costs of skateboarding and the advantages of a skateboard park.
- Student-selected items would ask students to 1) contrast a detail in one person's comments at the meeting to a detail in the second article or 2) identify causal connections suggested by the details in the two articles.
- Either student-constructed or student-selected item(s) would require the identification of the major arguments made for and against skateboards.

4. **Criteria for judging/evaluating responses:** The student-constructed responses would be judged on the basis of *accuracy, relevance*, and *sufficiency*. The evaluation would be based on whether the response:

   - demonstrated that the reader had (1) identified the major points and had (2) distinguished the major points from the supporting details;
   - demonstrated that the reader had understood the text's main ideas and the relevance of the details to support them; and
   - whether the response provided a sufficient amount of information to demonstrate that the reader had an understanding of the major points and the key details supporting each.

   The student-selected responses would be written with distractors that test these same criteria.

5. **Instructional suggestions:** The scoring of the responses should lead to an indication of how the teacher can reinforce reading strategies relevant to the activity. The kinds of instructional activities that would arise from the results of this assessment might include such things as:

   - Assigning students to read one or more texts on a topic rich in details but without graphic cues to the main and minor points. (One of these could be from a science or social studies text in use.) The teacher could have the students read and then, as a group, outline the topic/article on the chalkboard.
   - Teaching students how to synthesize and compare details across two or more texts with an emphasis on deciding which details are more important to particular perspectives posed by particular questions and tasks.
   - Demonstrating and having students read a narrative text and then make notes of details from it on the chalkboard. They would then discuss which of the details were the most important and create an outline of the main ideas and the supporting details.

Specifications for an Assessment Using a NARRATIVE Text

Purpose: ***Read to understand the theme of a story and how the author develops it.***

1. Specific reading purpose: The reader would be asked to identify his or her interpretation of the author's theme about friends who must part. The task is one of synthesizing the situation and comparing (contrasting) the attitudes of the two girls as those relate to the theme.

2. Reading text: The text would be a story about two girls who are very good friends. They have made plans about how they will be doing things together for years into the future. Then they learn that the mother of one of them has accepted a new job that requires that the family move to a different city. One of the girls talks about ways they can stay in touch and still be good friends: they can visit each other; they can get permission to call one another regularly; they can spend vacations at each other's house; they can stay in touch with e-mail; and they can write letters. The other girl is cynical and depressed about her friend's leaving. She insists they will not be able to remain really close friends…that their friendship will really be at an end. She is certain that her friend will make a

new best friend at her new school.

3. Assessment activities:

- A student-constructed item would ask the reader to explain the author's theme. What do you think the author is trying to tell us about friendships? Students would be asked to use the opinions of the girls to explain their interpretation of the author's theme.
- A student-constructed item would ask the student to write what the story says about growing up and about facing changes in friendships.
- Student-selected items would require the summarizing of each girl's attitude about the move as a part of the theme on friendship and growing up.
- Student-selected items would ask the student to select adjectives that describe the attitude of each girl about friendships.

4. Criteria for judging/evaluating responses: The student-constructed responses would be judged on the basis of *accuracy, relevance*, and *sufficiency*. The evaluation would be based on whether the response:

- demonstrated that the reader could articulate a reasonable story theme;
- demonstrated that the reader could explain how various elements of the story—in this case, character differences, contributed to the development of its theme; and
- whether the response could provide enough information to support the reader's interpretation of the story theme and the story elements directly related to it.

The student-selected responses would be written with distractors that tested these same criteria.

5. Instructional suggestions: The scoring of the responses should lead to an indication of how the teacher could reinforce reading strategies relevant to the activity. The kinds of instructional activities that would flow from the results of this assessment might include such things as:

- Teaching students to read to analyze how story characters often express a general attitude about life that is, in effect, the story theme.
- Consider various messages or general ideas that authors of different pieces of fiction are trying to get the reader to think about.
- Compare how different authors may have expressed their attitudes about the same general ideas or themes.

Purposeful Reading: Purpose-Specific Evaluative Criteria

As an illustration of how the three general evaluative criteria (*accuracy, relevance,* and *sufficiency*) could be applied to judge students' mastery of the purposeful reading skill, a set of purpose-specific evaluative criteria will conclude this analysis.

PURPOSE-SPECIFIC EVALUATIVE CRITERIA

FUNCTIONAL TEXTS

Purpose 1: ***Read to determine the relevance and importance of functional information.***

Accuracy	Relevance	Sufficiency
The reader demonstrates an accurate understanding of the information.	The reader recognizes whether key aspects of the content are relevant to a reader's informational needs.	The reader cites an adequate amount of information to support conclusions about the relevance and importance of the information read.

Purpose 2: ***Read to select and apply relevant information for a given task.***

Accuracy	Relevance	Sufficiency
The information selected from the text to be applied is used accurately. All the details applied are the same as in the text. For example, times, places, persons (contacts), events, etc. are correct according to the text.	The information selected is appropriate to the task. It is applied logically and validly to complete the task. For example, in following directions, the reader has included all the essential steps indicated in the text.	The reader selects enough information and details from the text to complete the task adequately and appropriately.

EXPOSITORY TEXTS

Purpose 1: ***Read to understand a text's major points and supporting details.***

Accuracy	Relevance	Sufficiency
The response demonstrates that the reader has 1) identified the major points and has 2) distinguished the major points from the supporting details.	The reader's response demonstrates that the reader has understood the text's main ideas and the relevance of the details to support each of them.	The reader provides an adequate amount of information to demonstrate that the reader has an understanding of the major points and the key details supporting each.

Purpose 2: ***Read to understand the text's organizational exposition and how that organization serves the writer's purpose.***		
Accuracy	**Relevance**	**Sufficiency**
The reader's response indicates an understanding of how the writer has organized the content, using features such as sequence and cause/effect, chronology, or categorization.	The reader's comprehension reflects an understanding of the organization of the text to support the writer's major concepts and purposes.	The reader cites an adequate number of details to support an understanding and reaction to the organization.
Purpose 3: ***Read to understand how the information in the text fits into broader topics and issues.***		
Accuracy	**Relevance**	**Sufficiency**
The reader uses the main concepts of the text and accurately relates them to commonly understood concepts about the world.	The reader uses the newly acquired information in a text to develop a growing comprehension of commonly understood concepts.	The reader can apply an adequate number of newly acquired details from the text to indicate an enlarged understanding of the commonly understood concepts they inform.
NARRATIVE TEXTS		
Purpose 1: ***Read to recognize and understand an author's development of character, setting, and mood as basic story elements.***		
Character Development		
Accuracy	**Relevance**	**Sufficiency**
The reader can understand characters in fiction using details from that story that accurately report what the characters do, say, and think; what other characters say about them; and what the writer says directly.	The reader understands how the depiction of major and minor characters is important to a story's plot and theme…or simply what about the description of characters makes them worth remembering.	The reader uses an adequate amount of details from a story to describe and discuss its characters and their roles.

Mood

Accuracy	Relevance	Sufficiency
The reader can identify the mood of a story by identifying details the author has used to establish it.	The reader can describe the importance of mood to the story's plot and theme.	The reader recognizes a number of techniques and details the author has used to establish one or more moods throughout a story.

Setting

Accuracy	Relevance	Sufficiency
The reader can accurately recall important details and descriptive techniques that distinguish the location and time in which a story takes place.	The reader can discuss a setting in terms of its relevance to the story's plot and theme.	The reader can report with adequate details each of the settings in which a story takes place.

Purpose 2: ***Read to understand how the plot of a story develops as a series of high points and/or how it can be depicted as a problem and its solution.***

Accuracy	Relevance	Sufficiency
The reader can correctly identify high points in a story's plot in relation to its highest point, the climax.	The reader can note how high points in a story's action build to its climax and/or can identify the presentation of a problem and its solution or resolution as basic to the story's plot structure.	The reader can adequately describe all of the significant events in a story and the sequence in which those events occur.

Purpose 3: ***Read to understand the theme of a story and how the author develops it.***

Accuracy	Relevance	Sufficiency
The reader can articulate a reasonable story theme.	The reader can explain how various elements of the story contribute to the development of its theme.	The reader can provide enough information to support his or her interpretation of the story's theme.

Appendix B
Problem-Focused Number Usage

Mary M. Lindquist
Fuller E. Callaway, Professor Emeritus
Columbus State University

Students learn mathematics in order to solve problems that require mathematical skills and concepts as well as to enable them to learn more mathematics to enlarge the repertoire of mathematical problems they can solve. Although the content standard described here focuses only on number, there are other mathematics content standards that must be included at each grade level.

One purpose of learning mathematics is to solve numerical problems; this requires acquiring those skills and concepts necessary to solve both routine and non-routine quantitative problems. It is possible, therefore, to conceive of a content standard focused on students' mastery of this important curricular outcome. The content standard of *Problem-Focused Number Usage* is centered on (1) the skills and concepts needed for the solution of grade-appropriate problems and (2) the application of those skills and concepts in the actual solution of such problems. Although number usage can be addressed in isolation, apart from a problem-solving context, this content standard deliberately treats number usage in relation to the solving of problems. By making the number usage skills and concepts to be taught by teachers explicitly relevant to the solution of routine and non-routine problems, students will be able to learn (and teachers will be able to teach) key number usage skills and concepts in a more meaningful manner.

Content Standard

Students can use numerical skills and concepts to solve routine and non-routine problems appropriate to their grade level.

Assessment Description for All Grade Levels

The curricular aim being assessed is the student's ability to *understand and use numerical concepts skills* associated with three types of numbers: *whole numbers, decimals, and fractions.* At each grade, the particular skills and concepts needed for problem-solution involving each of these types of numbers will be set forth as a grade-specific assessment description. Students' mastery of those skills and concepts can be assessed (1) separately, that is, without requiring their application in the solving of a particular problem, or (2) in order to solve *routine* problems (such as "Determine the price of three items given the price of each item.") or *non-routine* problems (such as those requiring the application of concepts and skills in an atypical

manner). If students are asked to solve a non-routine problem, the number usage skills needed for the problem's solution will be drawn from those that are at least one or two grades lower. Alternatively, a state may specify the use of calculators when students are solving problems with skills expected to be mastered at or above the grade level tested or when such a tool is essential. In problem-solving situations, students may need to use concepts and skills from other strands of mathematics such as measurement, geometry, or statistics.

Concept Assessment. Students will have mastered a concept if they understand that concept at an appropriate level. Students' understanding of a concept will be assessed according to whether they can

a) use the concept appropriately or

b) discriminate between proper/improper uses of the concept. In designated instances, a student's understanding of a concept will be assessed according to whether he or she can explain the concept appropriately.

Skill Assessment. Students display mastery of a skill when they can use it. Students' skill-mastery will be assessed according to whether they can (a) perform the skill at an appropriate level or (b) use the skill in a problem-solving situation.

Many types of items can be utilized, even on a paper-and-pencil test to assess students' understanding and use of mathematics. Students' responses to well designed selected-response items often can give some indication of their understanding of concepts and use of skills. Additional information can be obtained from items for which students must construct a response. Constructed-response items provide the opportunity to learn more about a student's understanding and use of mathematics. A generic rubric for evaluating students' responses to this type of item is given at the end of this appendix and is applied more specifically to the open-response sample items at each grade level.

Developing Grade Level Assessments

In developing assessments for the *problem-focused number usage* content standard at each grade level, the following list of questions may be helpful. This was used in developing the grade-specific assessment for grades 4 and 5 illustrated in this appendix. At the close of these grade-specific assessment descriptions, several illustrative sample test items are presented. These sample items are not intended to be exhaustive of the types of items that might be employed to measure students' mastery of this content standard. However, the items are intended to convey a general idea of the *upper limit of difficulty* that can be represented in the actual test's items.

NUMBER CONCEPTS	OPERATION CONCEPTS	NUMBER SKILLS	OPERATION SKILLS
What type of numbers? (e.g., whole numbers, fractions) What size of numbers? (e.g., 5-digit numbers, 2-place decimals) What concepts? (e.g., fractions describe parts of a whole, division, ratio; place-value; numbers can be represented in many ways)	What meanings of operations? (e.g., division is equal sharing) What relationships among operations? (e.g., division is the inverse of multiplication) What properties of numbers and operations? (e.g., addition is commutative but subtraction is not).	What skills with numbers? (e.g., counting by fives, hundreds, expressing numbers in different ways, ordering numbers)	What operations? (e.g., addition, subtraction, multiplication, division) What skills? (e.g., mental, paper-and pencil, calculator, estimation) What level of proficiency? (e.g., mastery of facts, efficient use of a division algorithm, or use of calculator to find powers of numbers)

The type of problems that students should be able to solve should be kept in mind when developing the answers to each of these questions. There should also be clear articulation from grade to grade so that concepts and skills are not repeated at the same level but developed more fully or proficiently. Additionally, there should not be large gaps in the development from grade to grade. For example, if the expectation in one grade is that students should solve problems involving the multiplication of fractions, then the conceptual background of fractions should have been previously developed.

Grade Four Assessment Description

The concepts and skills eligible for assessment in the fourth-grade tests are identified for each of the three types of numbers: *whole numbers*, *decimals*, and *fractions*. Test items call for a student to display mastery of those concepts or skills in isolation (as in an unadorned numerical problem) or as part of a solution of a verbally presented problem. Any skill or concept from previous grades may be used in problem solving.

Whole Numbers

Concepts

(a) place value including grouping by tens, hundreds, etc., the relationship of one place to another, and the value of a digit in any place
(b) representations of numbers with pictorial models or symbols
(c) algorithms (procedures) for each of the four operations (addition, subtraction, multiplication and division) at grade 4 level of use.

Skills

(d) read, write and order whole numbers
(e) estimate quantities
(f) add and subtract
(g) estimate sums and differences
(h) know multiplication facts at a fluent level
(i) multiply two-digit numbers
(j) divide any whole number by a one-digit number

Fractions

Concepts

(a) part-whole meaning of fractions (smaller and larger than 1)
(b) equivalent fractions

Skills

(c) read and write fractions
(d) use models to order, or to add and subtract simple fractions

Decimals

Concepts

(a) place value (two places)
(b) relation of decimals to whole numbers (extension of place-value of whole numbers), to fractions, and to money

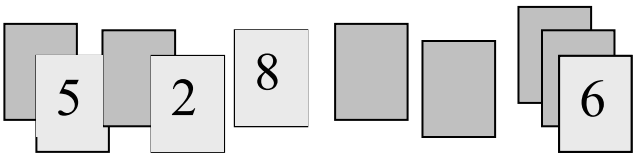
Skills

(c) read, write, and order decimals (two places)
(d) add and subtract decimals (two-digits)

Sample Items for Grade 4

1. This is an example of a non-routine problem in ***Constructed-response*** format that uses the concepts of place value. (***Whole Numbers*** a)

NUMBO GAME: A GAME FOR TWO



MATERIALS

A set of cards with each of the digits (0-9)
A three-place grid to record numbers for each player

RULES

Players take turns
Each player draws a card, looks at the number, and records the number in his or her three-place grid
A player should not change a number once it is recorded

WINNER

The player who makes the larger number

PROBLEM

Amy and Kyo are playing NUMBO.
Amy goes first and places the 2 in the ones place.
Kyo draws the 5 and places it in the tens place.
The results of this and the next draws are shown below.

	8	**2**

Amy

6	**5**	

Kyo

Amy says, "I can win only if I draw a 9."
Is she correct? Mark: YES OR NO
Explain you answer.

2. This is an example of a ***Selected-Response*** item focused on representing numbers symbolically. (***Whole Numbers*** b)

Which of the following is NOT a representation of 1803?

A. 1000 + 80 + 3
B. 1000 + 800 + 3
C. 18 hundreds + 3 ones
D. 1 thousand + 7 hundreds + 10 tens + 33.

3. This is an example of a ***Short-Answer Constructed-Response*** item focused on writing numbers. (Note: if this were used as a multiple-choice item, its focus would be on reading numbers.) (***Whole Numbers*** d)

   Write 305,970 in words: _____

4. This is another example of a ***Constructed-Response*** item; this one is focused on explaining procedures of subtracting. (***Whole Numbers*** c)

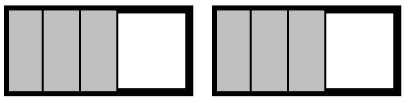
   Sam had developed a method to subtract numbers from hundreds such as 300 or 700. For example, Sam subtracted 397 from 600 by subtracting 400 and adding 3 to the answer. Although he did this in his head, he wrote this when he was explaining his method to a friend.

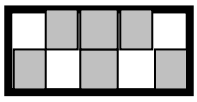
```
  600        600
- 397      - 400
             200
           + 3
             203
```

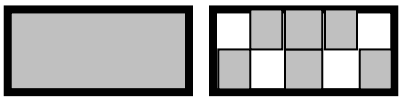
   A. Using Sam's method, show or tell how he would subtract 289 from 500.
   B. Susie said she could use Sam's method to subtract any numbers, not just from hundreds. Angela said it would be difficult for some numbers.
      a. Give an example of a pair of three-digit numbers to show what Susie meant. Show or tell how Susie would do the subtraction.
      b. Give an example of a pair of three-digit numbers that show why Angela thought it might be difficult. Explain the difficulty.

5. This is an example of a ***Selected-Response*** item focused on the part-whole meaning of fractions as well as equivalent fractions. (***Fractions*** a and b)

The large rectangle is the whole. Which of the following shaded areas represent $1\ ^{3}/_{5}$?

A. 

B. 

C. 

D. 

Answers to sample items

1. Level 4: Student explains or shows how to subtract 289 from 500 by first taking 300 from 500 and adds 11. Student also gives an example such as Susie's of subtracting a number from a non-hundred number such as 298 from 525 by subtracting 300 from 525 and adding 2 to the difference. To show that it may be difficult as Angela posited, the student gives an example in which the subtrahend is not close to a hundred may be used (e.g., 528-434) or an example which is much easier to do without this (e.g., 528 - 328).
   Level 3: Student adequately explains two of the three examples.
   Level 2: Student explains one of the three examples.
   Level 1: Student shows no understanding of the problem.

2. A.

.

3. Three hundred five thousand nine hundred seventy.

4. Level 4: NO. She could draw either a 7 or 9 since either 7 hundreds or 9 hundreds is greater than Kyo's 6 hundreds. (This could be said in different ways, but the value of the places or size of number must be mentioned.)

   Level 3: NO. She could draw either a 7 or a 9. (No explanation in this statement, but an understanding of the problem and knowing that 7 or 9 is greater than 6.

   YES. 9 hundreds would be more than Kyo's 6 hundreds.

   Level 2: NO. She could draw other numbers.

   Level 1: NO or YES with no explanation.

5. C

Grade Five Assessment Description

The concepts and skills eligible for assessment in the fifth-grade tests are identified for each of the three types of numbers: whole numbers, decimals, and fractions. Test items call for a student to display mastery of those concepts or skills in isolation (as in an unadorned numerical problem) or as part of a solution of a verbally presented problem.

Whole Numbers

Concepts

(a) properties of operations
(b) factors and multiples
(c) algorithms for multiplication and division at grade 5level of use.

Skills

(d) multiply and divide
(e) estimate products and quotients

Fractions

Concepts

(a) division meaning of fractions
(b) ratio meaning of fractions

Skills

(c) estimate sums and differences of fractions
(d) add and subtract simple fractions
(e) multiply a fraction by a whole number
(f) divide a fraction by a whole number

Decimals

Concepts

(a) place value (more than two-places)
(b) equivalency of decimals and fractions

Skills

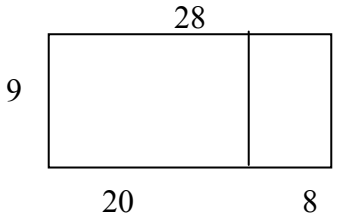
(c) add and subtract decimals
(d) multiply and divide a decimal by a whole number
(e) estimate sums, differences, products, and quotients

Sample Items for Grade 5

1. This is an example of a ***Yes-No Selected-Response*** item focusing on properties of operations. (***Whole Numbers*** a)

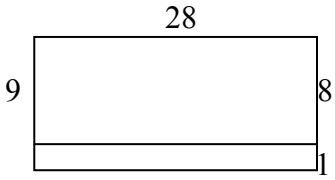
   Which of these shows that 9 X 28 = (9 X 20) + (9 X 8). Circle YES or NO for each of the following. (The pictures are not drawn to scale.)

A.



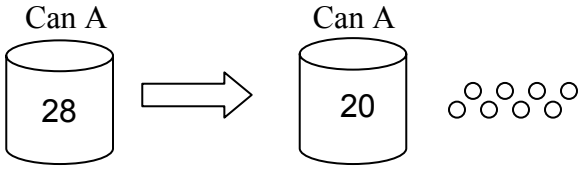
YES or NO

B.



YES or NO

C. Each of 9 cans holds 28 marbles. Eight marbles are moved from each can as shown for Can A below.



YES or NO

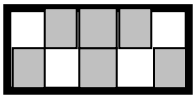
2. This is an example of a routine multiplication problem given in ***Short-Answer Constructed-Response Format***. (***Whole Numbers*** d)

   Shana's class measured the distance of a hall in their school and found that it was 45 yards. How many inches is that? (1 yard = 36 inches).

3. This is an example of a ***Multiple-Choice*** item focusing on the meaning of fractions. (***Fractions*** a and b)

   Which of the following would NOT be represented by 3/5?

   A. The shaded part of the rectangle shown below.



B. The amount of pie each of three people would receive if 5 pies were shared equally.

C. "Three of a family's five children are boys"

D. "The value of 3 nickels each worth 5 cents"

4. This is a ***Constructed-Response*** item that focuses on the equivalency of decimals and fractions. (***Decimals*** b).

   Jeremy says that 1/3 = 0.33. Jake says that 0.33 is only approximately equal to 1/3. Explain, using words, numbers, or pictures what Jake means.

5. This is an example of a non-routine problem in CONSTRUCTED-RESPONSE format that utilizes whole-number skills from lower grades but requires a deeper understanding of the situation than would be expected at lower grades.

   Zach is beginning an exercise program. The instructor says he wants Zach to move up the levels until he is doing 20 exercises each day. Here is the schedule of the number of days and the number of exercises at each level.

Level 1	3 days	2 exercises each day
Level 2	6 days	4 exercises each day
Level 3	9 days	6 exercises each day

   Assume that this pattern of 3 more days at each level and 2 more exercises at each level continues. How many days will Zach have to exercise before reaching the goal of 20 exercises each day?

Answers for sample items

1. A. YES; B NO; C YES

2. 1620 (inches)

3. D.

4. Level 4: Argues that 3 one-thirds is a 1, but 3 x (0.33) is 0.99 so it is almost 1 or approximately 1. Thus, 1/3 and 0.33 are approximately equal. OR Shows a square of 100 hundredths and shows that 0.33 is not quite 1/3. OR Explains that if 1 is divided by 3, there is always a remainder so it would keep going forever (0.33333…), but 0.33 is 0.330.
   Level 3: Gives a partial explanation of one of the above, but it is not complete.
   Level 2: My calculator says it is 0.3333333 (or number of places in student's calculator).
   Level 1: They are equal some of the time, or some such answer.

5. Level 4: Use table, listing or other method to ascertain that at Level 9 Zach would be doing 18 exercises a day for 27 days. He then has exercised (3 + 6 +…+ 24 + 27) days or 135 days.
Level 3: Show understanding of the problem and the pattern of numbers, but either includes Level 10 (the days he did 20 exercises in the total) or makes an arithmetical area in summing the days.
Level 2: Arrives at the wrong Level, or shows some understanding of the pattern but says that will be after 27 days (or 30 days) rather than summing the days.
Level 1: Shows no evidence of understanding the problem.

Generic Rubric

Open-response items should be interpreted in terms of the amount of evidence that a student provides as to his or her understanding and use of concepts and skills in solving each problem. The following rubric has been used in interpreting the sample items included in this appendix. The levels have been numbered for convenience only. They do not have to correspond to the scoring points awarded in an assessment situation.

Level 4. Complete evidence. Student response shows thorough understanding and use of the concepts and skills involved in solving the problem.

Level 3. Adequate evidence. Student response shows understanding of the concepts and proficient use of skills involved in solving the problem. There may be some minor misunderstanding of one of the concepts or a minor inaccuracy in a procedure.

Level 2. Limited evidence. Student response shows either a lack of understanding of the concepts or uses skills inaccurately; however, there is evidence of some understanding or some use of skills.

Level 1. No evidence. Student response shows little or no evidence of understanding and skills needed, or a grasp of the problem.