CICSD

High-Stakes Testing: The Background Behind Testing-Based Educational Reforms and Implications for Practice by Speech-Language Pathologists

Monica Gordon Pershey
Cleveland State University, OH

ABSTRACT: Some current educational reforms (U.S. Department of Education, 2002) focus on preparing students for mandated testing. The Individuals with Disabilities Education Act (IDEA, 1997) requires education agencies to provide students with disabilities with the opportunity to participate in testing. Speech-language pathologists face the challenge of helping students with language and learning disabilities and associated cognitive differences prepare for test requirements.

The purpose of this article is to provide background information on mandated testing, including (a) differences between testing students' minimum competency versus curriculum mastery, (b) differences between norm-referenced and criterion-referenced tests, (c) potential limitations of basing accountability decisions on the results of testing that is administered once, (d) implications of revising curriculum to reflect material to be tested, and (d) testing concerns for students from underserved populations or with learning challenges. Approaches for involving speech-language pathologists in curriculum-based speech-language interventions that prepare students for testing are proposed.

KEY WORDS: assessment, testing, speech-language pathology, education reform

Some current educational reforms are concerned with establishing policies and strategies that will ensure the accountability of educational agencies (i.e., school districts, state boards of education, individual public or private schools). One accountability strategy is the use of mandated testing (U.S. Department of Education, 2002). The term "high-stakes testing" is applied when the required testing has important consequences for students, schools, districts, personnel, and/or communities. As Kohn (2000, p. 5) stated, high-stakes testing relates to circumstances where "a test is made to 'count'—in terms of being the basis for promoting or retaining students, for funding or closing down schools."

The National Center on Educational Outcomes (NCEO) (2001) reported that poor performance on high-stakes tests may result in students not receiving a high school diploma or not being promoted to the next grade. States may sanction schools or districts by assigning negative labels, such as "low-performing," "at risk," or "under academic watch." Educational and administrative personnel at all levels may be removed from their jobs. Communities may suffer when newspapers publicize low test scores. On the other hand, in schools and districts where test performance is strong, states may award the schools extra funding, positions and programs may be created, and communities may prosper.

The purpose of this article is to provide background information on high-stakes testing that may be useful to speech-language pathologists (SLPs) who serve school-age children and teens, specifically those whose language and learning disabilities and associated cognitive differences place them at risk for diminished performance on tests (American Speech-Language-Hearing Association [ASHA],

1092-5171/03/3001-0047

2000). The first section of this article will discuss some of the characteristics of tests, describe aspects of the rationale behind testing, and explore some of the potential impact of high-stakes testing. This article is not meant to be an indictment of testing; rather, the intent is to help SLPs become familiar with some of the issues and concerns attendant to testing. SLPs' participation in preparing students to perform on tests is a necessity for students' success and is a professional obligation (ASHA, 2000). The Individuals with Disabilities Education Act (IDEA, 1997) charges school personnel with preparing students with special needs to meet contextual demands, which may include high-stakes tests. To participate, SLPs need to be familiar with the types of tests their students are taking, the rationale behind such tests, and the credibility, merit, applicability, comparability, and implications of test scores.

Next, specific considerations for test selection and use will be addressed in this article, including (a) the important differences between testing students' minimum competency versus their curriculum mastery, and (b) the advantages and disadvantages of norm-referenced and criterion-referenced tests. Additionally, given the tremendous significance that test scores have in many schools, districts, and communities (Kohn, 2000), this article will describe some of the potential limitations of the theory of performance acquisition (Glaser, 1994), that is, basing accountability decisions, and at times ensuing school reforms, on the results of onetime testing that purports to reveal what students have learned from exposure to curriculum (Jones, 1997; Winfield, 1990). With increased awareness of some of the limitations inherent in onetime testing, SLPs can be a part of educational teams that evaluate students' onetime test scores and that develop allied means of monitoring students' progress that are used in conjunction with test scores. SLPs are particularly well able to provide ongoing clinical data on students' performance that may help explain why students obtain certain scores.

Furthermore, part of the discussion about the suitability, utility, and validity of high-stakes testing centers on whether testing coincides with curriculum. SLPs are undertaking to fulfill curriculum-prescribed roles within the greater educative community (Ehren, 2000). This article will discuss how, in some cases, high-stakes testing reflects the curriculum and instruction that students have experienced, whereas in other settings, tests do not reflect the curriculum per se. In cases where testing coincides with curriculum, meeting curriculum standards by implementing curriculum-based speech-language interventions may facilitate preparing students for high-stakes testing. Where testing does not coincide with curriculum, students and SLPs face the challenge of meeting dual contextual demands—curricular objectives and test requirements.

Additionally, some concerns relating to testing students from underserved populations will be addressed. Finally, this article will consider approaches that SLPs can use to prepare students with language and learning disabilities and associated cognitive differences for high-stakes tests.

THE ROLE AND IMPACT OF HIGH-STAKES TESTING AND THE RATIONALE FOR ITS USE

The effort to establish tests of academic performance overlaps the current school accountability movement, which promotes the expectation that schools and educators should be held responsible for students' progress (National Education Association [NEA], 2001a, 2001b). In some settings, accountability efforts may serve to increase the stakes for students and teachers and promote actual attainment of higher educational standards (Falk, 2000). In other settings, it may be that the perceived or purported stakes are higher but instructional practices do not improve commensurately (Raivetz, 1992; Webb, 1995; Winfield, 1990). The presumption behind high-stakes testing is that if students, educators, schools, and districts are held accountable for test performance, individuals will be motivated to do their best and educational programs will improve (NCEO, 2001). A fairly common working assumption (that may be difficult to research) is that the presence or absence of school accountability can be inferred from test scores (Kohn, 2000). It might be said that test scores serve as proxies for accountability, in that it may be maintained that impressive test scores demonstrate accountability.

Although the promise of testing might logically be perceived as a powerful impetus for learning (Kohn, 2000; Tapper, 1997), a search of the literature on high-stakes testing reveals that it has not been stated definitively that the prospect of testing motivates school learning (Maddaus, 1988; Meisels, 1989), that test preparation fosters school learning (Allington & McGill-Franzen, 1992; Haladyna, Haas, & Allison, 1998), or that test results influence subsequent school learning (Shepard & Dougherty, 1991). The diversity of local and state testing policies makes it hard to characterize whether learning outcomes are directly enhanced for students who must perform on high-stakes tests (Koretz, 1991). Nevertheless, the presumption is that strong scores establish that test-takers' schooling has caused them to possess substantial skills and knowledge.

As a case in point, attributing reading achievement to the prospect of testing is quite controversial. According to the National Assessment of Educational Progress (NAEP, 1998, 1999b), improvement in reading scores nationally cannot be attributed to the implementation of competency-based testing programs. NAEP (1998; 1999b) studies of more than 10,000 fourth graders nationally found no effects for testing, either positive or negative, on reading achievement. Although positive effects for reading achievement were seen in grades eight and eleven for schools that test, much of this gain can be attributed to remedial and supplementary reading instructional programs geared toward boosting test performance, not to the fact that students were better prepared because of the promise of testing.

Nevertheless, for the general public, student performance on high-stakes tests may be the primary indicator of the standing of a school or district (NEA, 2001a). High-stakes tests increasingly have become the standard by which pupil competency and school accountability are assessed (Kane,

1994; Lanese, 1992). Tests have the potential to become the driving force behind educational decision-making (e.g., curriculum design, textbook adoption, scheduling, student assignment to classes or groups) (Kohn, 2000; Tapper, 1997). In a Tarrance Group/Quinlan Greenberg Research poll of 1,000 voters, 78% approved of annual testing of student performance in grades three to eight and 72% approved of annual testing in all grades (NEA, 2001a). Although only 16% of those polled believed that test scores alone are the best indicator of a school's performance, news media, politicians, and realtors freely announce students' scores and relate them to their specific agendas (Haladyna et al., 1998).

Given that there is perhaps insufficient research to claim that testing motivates performance (Heubert & Hauser, 1999; "High Stakes Testing," 1999; Linn, 1994; Phelps, 1999), testing might be established as a mechanism for assessing pupil progress and determining a student's standing relative to other students (U.S. Department of Education, 2002). It may or may not be possible to infer pupil progress or standing from a high-stakes measure (Popham, 1999). Many investigators argue that the high-stakes tests currently available are not valid grounds for academic tracking (i.e., "tracks" are designations of students' placement based on ability or curriculum focus) or grade retention. Nor, as instruments, are they mutually comparable, statistically secure, developmentally appropriate, culturally relevant, or intellectually meaningful (Airasian, 1988; American Educational Research Association [AERA], 2000; Barton, 1999; Commission on Behavioral and Social Sciences and Education, 1998; Denoyer & White, 1992; Etsey, 1997; Haladyna et al., 1998; Heubert & Hauser, 1999; "High Stakes Testing," 1999; International Reading Association [IRA], 1999; Kane, 1994; Lanese, 1992; Maddaus, 1988; Manning, Lucking, & MacDonald, 1995; Phelps, 1999; Stephens, 2000; Stroud, 1995; Webb, 1995).

High-stakes testing appears to be an issue that provokes strong feelings on both the pro (Barton, 1999; Linn, 1994) and the con (Airasian, 1988; AERA, 2000; Haladyna et al., 1998; Maddaus, 1988) sides of the debate. Notably, even among those on either side who are ardent, many people may not be fully informed about the purposes for testing and the potential for inaccuracy when accountability decisions are based on the results of summative tests that are administered only once (AERA, 2000). Individuals may not be aware of concomitant issues, for example, that educational inequities may be borne out in testing performance (Manning et al., 1995; Winfield, 1990) or that standardized testing may not correspond to the curriculum that students have been taught (Barkley, 2001; Barton, 1999; Heubert & Hauser, 1999; IRA, 1999). It is critical that SLPs be aware that the existence of these issues in some settings does not result automatically in misapplication of testing in all settings. Education agencies can select testing that is useful and can engage in testing practices that are closely related to student learning. It is important for SLPs to be informed about specific considerations for test selection and use and to be able to classify the types of tests available and differentiate their purposes.

DIFFERENTIATING TESTS OF MINIMUM COMPETENCY FROM TESTS OF CURRICULUM MASTERY: THEIR USE AS HIGH-STAKES INSTRUMENTS

Education agencies may administer high-stakes tests for either of two main reasons: to obtain evidence of (a) students' minimum competency or of (b) curriculum mastery (Bond, 1995). Testing may be conducted at several grade levels, from elementary through high school, in an effort to increase reliability through repeated measures and to obtain cross-sectional and/or longitudinal data on either minimum competency or curriculum mastery (AERA, 2000; Winfield, 1990).

In the 1990s, many education agencies adopted assessments of minimum competency and required them for a transition, such as high school graduation or grade promotion (Barton & Coley, 1994; Lanese, 1992; Meisels, 1989; Raivetz, 1992; Robinson & Moore, 1992). Minimum competency tests typically bear little correspondence to curriculum content for the grade level at which they are administered (Linn, 1994). Rather, the tests intend to demonstrate achievement of rudimentary knowledge and skills. There is seldom an attempt to evidence that the required knowledge and skills were acquired at school or, more particularly, at grade level (Etsey, 1997; Gallagher, 1993; Kane, 1994).

A test of minimum competency may be of great importance to challenged students (Allington & McGill-Franzen, 1992) and may reflect some of the remedial work that SLPs help students accomplish, but this sort of test is inconsequential to most average students—those who have attained basic skills but who are contending with the curriculum (Linn, 1994). For successful students, a basic skills test would lack curricular and instructional validity (Barton, 1999). Therefore, education agencies may opt for tests that purport to measure curriculum mastery (Airasian, 1988; Barton, 1999; Barton & Coley, 1994). If accountability is a goal, tests that reveal whether students master curriculum may substantiate the efforts of students, teachers, and SLPs (Kane, 1994; McGee, 1997; Robinson & Moore, 1992; Texas Reading Initiative, 2002; U.S. Department of Education, 2002).

Tests of curriculum mastery are in line with "opportunity to learn" (OTL) testing (Linn, 1994). Content validity is demonstrated when tests correspond with what is taught (i.e., the content and skills that students had the opportunity to learn in school) (Kane, 1994; Popham, 1994, 1999; Pottle, 2001). In some settings, scores inform students, parents, and other stakeholders of the students' progress toward targeted learning outcomes (Barkley, 2001) and may be used to help teachers and SLPs design subsequent instruction to redress weaknesses (Falk, 2000; Jones, 1997; Popham, 1999; Shepard & Dougherty, 1991).

High-stakes tests of both minimum competency and curriculum mastery are summative, that is, they are given in order to summarize students' final performance, be it at completion of a grade level, at graduation, or over the interval since testing was last given (AERA, 2000; Barton,

1999; Heubert & Hauser, 1999). This stands in contrast with formative assessment, where school personnel gather frequent data on student performance, preferably using a variety of measures, that may include testing, collecting student work samples, conferencing with students, and keeping anecdotal notes of classroom observations—practices that may be familiar to many SLPs (ASHA, 2000; Nelson, 1998).

NORM-REFERENCED AND CRITERION-REFERENCED MEASURES

Whether a test is designed to reveal minimum competency or curriculum mastery, education agencies rely on two types of summative tests of achievement: norm-referenced and criterion-referenced measures. According to Bond (1995), norm-referenced standardized tests are based on data gathered from a broad cross-section of learners. Questions are prepared that will produce response variance. A student's score is compared to the scores obtained by the students in the standardization sample. The objectives of norm-referenced testing are to see where a student scores given a range of possible scores and to rank students from high to low achieving (Bond, 1995). Norm-referenced tests usually establish that age and ability are covarying gradients (i.e., it is presumed that older students have greater knowledge and skills; as age increases, so does performance). SLPs frequently administer standardized tests, especially when testing language development and competence.

Districts often conduct large-scale administrations of well-known norm-referenced achievement tests, such as the Iowa Tests of Basic Skills (ITBS; Hoover, Hieronymus, Frisbie, & Dunbar, 1996), the Stanford Achievement Test (1996), the Metropolitan Achievement Test (MAT; Balow, Farr, & Hogan, 1992), the California Achievement Tests (CAT; CTB/McGraw-Hill, 1992), and the TerraNova (CTB/McGraw-Hill, 1998). Test authors strive to select content that matches grade-level learning outcomes (Bond, 1995), but this is a difficult proposition. Even if only one curricular area, for example, language arts, is considered, state curriculum standards vary widely (Stotsky, 1997). It would be very difficult for questions to correspond to 50 state curricula, let alone to thousands of district courses of study, plus the objectives addressed by the scope and sequences of the large number of textbook series in use, as well as school building-based mandates (Bobbett, 1993; Commission on Behavioral and Social Sciences and Education, 1998; Kane, 1994; Koretz, 1991; McGee, 1997; NAEP, 1999b; Stotsky, 1997).

Norm-referenced achievement tests are designed to measure how well a student's school learning compares to that of his or her peers (Bond, 1995). These achievement tests address a very different construct from the norm-referenced abilities or developmental tests that SLPs commonly administer, which are intended to assess how well a child's maturation in a particular area compares to that of his or her peers (Nelson, 1998). That is, in order to diagnose disorder or delay, testing helps determine whether the child's development is commensurate with that of a peer group. Achievement and abilities tests have in common the fact that performance is influenced by the range of experiences and stimulation that a test taker has had, cognitive and language skills, and comfort during test taking (Nelson, 1998; Robinson & Moore, 1992).

Barton and Coley (1994) examined NAEP and General Accounting Office summaries of state testing programs and found that 34 states use norm-referenced tests (some states use more than one test and/or also use criterion-referenced tests). Questions of comparative reliability and validity arise because there are so many tests from which to choose, all of which use different normative samples and test different content and skills (Etsey, 1997; Koretz, 1991; Phelps, 1999; Pottle, 2001). Each test may be reliable and valid in itself, and several studies have shown that some of these different measures do correlate (both positively and negatively, when examining various subtests at various administrations) (Pershey, 2001; Robinson & Moore, 1992). However, Stroud (1995) argued that even high correlations do not constitute concurrent validity, which ought to come from an accumulation of results from different sources. This can be accomplished potentially by the use of consistent testing across educational agencies. Current efforts are underway to prepare national testing based on academic content area standards that will allow education agencies across the nation to compare children on a common scale (U.S. Department of Education, 2002).

As an alternative to norm-referenced tests, an education agency or state government may commission the construction of criterion-referenced tests that correspond to curricular areas (Popham, 1994). Criterion-referenced achievement tests attempt to measure acquired, cumulative knowledge and skills as demonstrated by a behavioral performance (Glaser, 1994). A learner must answer a certain number of items correctly in order to meet a criterion set for passing the test (Bond, 1995). This type of test may assess minimum competency or be an OTL test given at intervals, traditionally in mid-elementary school, during middle school, and at least once in high school (Barton & Coley, 1994; Lanese, 1992; Meisels, 1989; Raivetz, 1992; Robinson & Moore, 1992). These tests may be valid if they match the curriculum and assess what a pupil has had the opportunity to learn in school (Linn, 1994). Barton and Coley (1994) reported that 34 states use criterion-referenced tests (some use multiple tests and/or both norm-referenced and criterion-referenced tests). The Commission on Behavioral and Social Sciences and Education (1998) determined that state, commercial, and NAEP criterion-referenced assessments vary too considerably in the thought processes they require and in their content, emphasis, and types and difficulty of questions to actually be comparable. The Commission could not devise a single equivalency or linking scale, nor could they propose a method by which scores could be converted to NAEP equivalents. Similar attempts to identify equivalencies in how test scores might be interpreted have revealed an allied concern, namely that there is no uniform way for depicting progress or proficiency across settings (NEA, 2001b). Test scores designed to compare individuals appear to be potentially unsuitable

measures for comparing communities. This is especially so when comparing communities that differ in affluence, resources, or racial or ethnic composition, or in cases where pupils in different communities took different criterion-referenced tests (Bobbett, 1993; Gallagher, 1993; Manning et al., 1995; Webb, 1995).

In summary, norm-referenced standardized tests may not reflect local curriculum (Gallagher, 1993; Webb, 1995) and criterion-referenced tests may not allow for comparisons of students across settings (Barton & Coley, 1994; Lanese, 1992; McGee, 1997; Robinson & Moore, 1992; Stroud, 1995). These limitations mean that educational setting is a variable that is not taken into account when test scores alone are reported (Barton & Coley, 1994). A criterion-referenced state-mandated test may be, for all practical purposes, a test of minimal achievement in an affluent, high-achieving school that teaches concepts and skills that are far more complex than what is prescribed by the state curricula. But the same test might be very rigorous for students in challenged schools where academic demands may not be elevated beyond state minimum requirements (Lanese, 1992; McGee, 1997; Pershey, 2001; Robinson & Moore, 1992). Both norm-referenced and criterion-referenced achievement tests purport to measure how well a pupil's school learning compares to that of his or her peers, but this is difficult to do when the variable of school environment cannot be controlled (AERA, 2000; Bobbett, 1993; Gallagher, 1993; Ladson-Billings, 1994; Linn, 1994; Sheridan, 2000; Webb, 1995).

It also might be that norm-referenced testing of curricular concepts and skills could present significant challenges to students with language and learning disabilities and associated cognitive differences who receive the services of SLPs. Recall that students who are served by SLPs generally qualify for services based on testing that has determined that their performance is not commensurate with a standardization peer group. Even so, fundamental to both minimum competency and OTL achievement tests, whether they are norm- or criterion-referenced, is the theory of performance acquisition, that is, the supposition that students respond as they do because of exposure to content acquired at school (Airasian, 1988; Gallagher, 1993; Glaser, 1994; Kane, 1994; Popham, 1999; Tapper, 1997).

ASSESSMENT THROUGH HIGH-STAKES TESTS: THE THEORY OF PERFORMANCE ACQUISITION

As has been stated, underlying achievement testing is the theory of performance acquisition, that is, the assumption that students respond as they do ostensibly because of exposure to content acquired at school (Glaser, 1994). However, the theory's testable hypothesis would be whether test performance (dependent variable) would be a direct measure of learning that takes place in the context of school exposure (independent variable) (Airasian, 1988; Glaser, 1994; Tapper, 1997). Many circumstances limit the possibility of measuring the influence of this independent variable on this dependent variable. For a variety of reasons, test performance may be misleading and may not reliably reflect what has been learned from school exposure. As alluded to by many researchers (Airasian, 1988; Barkley, 2001; Bobbett, 1993; Denoyer & White, 1992; Gallagher, 1993; Gaskins, 1998; Glaser, 1994; Haladyna et al., 1998; IRA, 1999; Jones, 1997; Kohn, 2000; McGee, 1997; Popham, 1999; Pottle, 2001; Tapper, 1997), it generally is not possible to say that test performance is a clear and true indicator of what students were exposed to through schooling. A composite of the limitations of the theory of performance acquisition, as raised by these authors, is provided below. These limitations invoke consideration of a null hypothesis, that test performance (dependent variable) would not be a direct measure of learning that takes place in the context of school exposure (independent variable) (and perhaps might be a measure of the influence of other variables).

Limitations

The first limitation of the theory of performance acquisition is that test-taking circumstances may prevent students from demonstrating learning acquired via school exposure. Potentially, learning may not be revealed because (a) students may fail to apply themselves during testing; (b) learners are often required to work under time pressures—given more time, students' scores may be higher; and (c) test anxiety may lower scores.

A second limitation of the theory is that high-stakes testing is rarely administered in a pretesting-teaching-posttesting sequence. Rather, summative testing shows only students' end states. It is not known how far students were from this end state before instruction began. Again, given that some education agencies use tests that do not correspond to curriculum, it may not be evident whether students' summative performance is related to whether students acquired content and skills that were taught by teachers and SLPs. Indeed, where tests correspond to curriculum, teachers and SLPs may not have even taught the prescribed curriculum, but students could still have scored well on high-stakes testing if their performance is based on their prior or personal knowledge. Accountability may be compromised by the fact that a legitimate independent variable (whether students are exposed to the prescribed curriculum) is not measured directly—in the one case because testing and curriculum do not correspond and in the other case because students may have scored just as well on pretesting, before receiving any teacher or SLP input.

A third limitation is that it is problematic to assume that school exposure is equally meaningful to all students. Temporary or abiding issues of motivation, comprehension, health and wellness, emotional distress, and a host of other factors can render school exposure more or less meaningful.

A fourth limitation is that because of differences among learners, the theoretical claim that a test can be used as a measure of learning that has resulted from exposure to an instructional program is not realistic. If, for example, the instructional program was somehow inappropriate instruction for the learners, then learners also are taking an inapplicable test.

Fifth, high-stakes tests allege to measure whether students can or cannot demonstrate specific knowledge or skills. Performance sampled on demand, using one or very few items, is the basis for an inference about what test takers know about a given domain (their overall exposure and retention). Important judgements about students' knowledge or skills are made given only a small amount of evidence.

Sixth, high-stakes test scores may reveal nothing about how much of a given capability students possess, how far each student may be from mastery of a concept or skill, and what knowledge or behaviors are missing. Seventh, students, as individuals, may not have the resources to respond as well to one item type as to another item type (Popham, 1994), but scores typically do not describe or differentiate what types of responses have been elicited on a test or subtests. An eighth limitation is that tests reveal learning that students acquired through classroom instruction, but also learning that occurred out of school; this may reflect the socioeconomic status of families and/or communities (Bobbett, 1993; Denoyer & White, 1992; Gallagher, 1993; Lanese, 1992; McGee, 1997; Meisels, 1989; NAEP, 1999a; Popham, 1999; Raivetz, 1992). A ninth and final limitation is that other factors, such as the adequacy of the schooling students have received, lack of attendance in school, greater or lesser school resources, and individual special needs, are not taken into account when tests are scored (NEA, 2001b).

Criterion-referenced tests present additional limitations to the theory of performance acquisition that are not encountered with norm-referenced tests. First, test makers must determine a cutoff or threshold for performing up to a criterion (Popham, 1994). A debate persists on what constitutes mastery-level performance on a test item, a subtest, or a test as a whole. According to Kane (1994) and Webb (1995), there is no agreement among psychometricians as to what performance standards ought to be or how a passing criterion on any given instrument is determined. A related second concern, then, is defining who is expert enough to set the standards for passing parts of the test or the entire test (Bond, 1995). Third, without a range of developmental norms that indicate typical performance on test items by students of various ages, it is not readily apparent whether the test content is suitable for the age group of students taking the test (Bond, 1995; Kohn, 2000).

In summary, the limitations of the theory of performance acquisition may impede faith in the reliability and representativeness of students' performance on a summative test that is administered only once in a school year (Etsey, 1997; IRA, 1999). It is important to recognize once again that high-stakes test results may not have reflected school curriculum and teachers' and SLPs' instructional practices (Airasian, 1988; Barkley, 2001). This is an often-cited limitation of norm-referenced standardized achievement tests (Kohn, 2000). To bridge this gap, criterion-referenced tests are developed to reflect curriculum, usually state-mandated curriculum (Linn, 1994). It also is possible that curriculum and instruction are redeveloped after a criterion-referenced test is constructed in an effort to align school practices with the knowledge and skills assessed by the high-stakes measure (Gallagher, 1993; "Texas Reading Initiative," 2002) and to allow SLPs to plan services that coincide with test demands.

HIGH-STAKES TESTS AND ISSUES RELATED TO CONVERGENCE WITH CURRICULUM AND INSTRUCTION

All 50 states test summative performance and 49 states have curriculum standards in place, but there are no clear data on how many states have aligned their testing programs successfully with their curriculum standards (NEA, 2001b). In some settings, critical decisions may have been based on the results of tests that do not reflect the schooling that students have experienced (Heubert & Hauser, 1999; Raivetz, 1992). Education agencies that are mindful of accountability policies might react to a lack of convergence by restructuring curricula and instructional practices to align more closely with the knowledge and skills required to score well on high-stakes tests (as conducted by the Texas Reading Initiative [2002]; see also "Texas Reading Success Network: Year 1," 2002).

The prospect of testing may impact curricular and instructional decisions, such as what is taught and how ongoing classroom performance is assessed. Alignment of curriculum and assessment is regarded as an efficient and valid process ("Texas Reading Initiative," 2002). The result might be to narrow the focus of instruction and ongoing assessments to correspond to the content and skills that are being tested. Ongoing classroom assessment may emphasize repeated practice of the types of questions found on the test. In some cases, anticipation of test content may orient instruction toward the practice of skills or memorization of factual information (Kohn, 2000). Conversely, many tests require learners to demonstrate that they can apply process skills (e.g., summarization, finding paragraph details, writing a set of directions to complete a task). The task is to carry over and apply the process skills that were practiced in class in order to operate on unfamiliar test content (Barkley, 2001; Pershey, 2001; Pottle, 2001).

Although curriculum and test alignment is efficient and valid, in less than ideal circumstances, test preparation may happen at the expense of teaching practices that promote critical thinking, cultural congruence, developmentally appropriate experiential learning, thematic teaching, and language development across the curriculum (Gordon & Reese, 1997; Jones, 1997; Ladson-Billings, 1994; Popham, 1994; Shepard & Dougherty, 1991). Ironically, the practices that are abandoned are exactly those practices that enhance schema development and are relevant to teaching students with language and cognitive deficits (Nelson, 1998) or with limited experiential backgrounds. Reduction of varied academic opportunities in order to meet accountability requirements is potentially likely to occur in schools where students are most in need of vivid learning experiences—schools that serve students who are economically challenged and/or of minority status (Kohn, 2000).

HIGH-STAKES TESTS AND STUDENTS WHO ARE UNDERSERVED AND/OR HAVE SPECIAL NEEDS

Questions arise concerning the appropriateness of standardized tests for various groups of students. Testing students with special needs, either diagnosed or undiagnosed, as well as those of racial, ethnic, or language minority status and those at risk for academic failure because of a lack of resources or a lack of opportunities is itself controversial (Bobbett, 1993; Gallagher, 1993; Kohn, 2000; Ladson-Billings, 1994; McGee, 1997; Meisels, 1989; Popham, 1999; Raivetz, 1992; Roderick, 1995; Winfield, 1990). Sheridan (2000) referred to a study sponsored by the Ohio Governor's Commission for Student Success that claimed that it is not reasonable, for example, to expect a teacher to bring a student entering fourth grade with second-grade capabilities up to fourth-grade standards in order to pass a test. Regardless of the strength of the teacher's efforts, a fourth-grade test probably would not be the appropriate test of achievement to give to such a child.

When students cannot pass a high-stakes test, a variety of methods of redress are considered, including grade retention; assessment for qualification for special education services; regular education initiatives, such as supplemental reading instruction; and parent-initiated supports, such as private tutoring (NAEP, 1998, 1999a, 1999b). Grade retention, largely resulting from failure to pass a high-stakes test, has received a great deal of scrutiny (Allington & McGill-Franzen, 1992; "High Stakes Testing," 1999; National Educational Research Policy and Priorities Board, 1999; Roderick, 1995; Westbury, 1994). According to these reports, and noting that these data reflect some cases where the determination to retain in grade was made independent of test scores, more than 30% of 14-year-olds are enrolled below grade level. It is not certain that test failure increases the prevalence of overage students in grade, but the number of overage students in grade level has risen by 40% in the past two decades. Some 20% of children experience retention by grade eight, but 52% of African American 14-year-old males and 30% of African American 14-year-old females have been retained.

Several studies have shown that minority and low-income students are more likely than non-minority and high-income students to fail high-stakes tests and that remediation for those who fail is less likely to be effective (NAEP, 1998, 1999a, 1999b, 2000; Winfield, 1990). The National Educational Research Policy and Priorities Board (1999) commented that "Retention…and segregated special education programs that stratify opportunities to learn by race, class, and gender do not result in high achievement for all students" (p. iv). Repeated test failure and remediation attempts can have a cumulative negative impact on students (Winfield, 1990).

Issues may arise when high-stakes test preparation determines too great a proportion of the curricular and instructional offerings given to students. Preparing students for high-stakes testing involves asking learners to store quantities of knowledge and skills (Kohn, 2000; Popham, 1994). A test-driven curriculum may alienate pupils who are already at risk for school disengagement and leave educators and SLPs little opportunity to design curriculum, instruction, and assessment that might be more captivating (Ladson-Billings, 1994; Manning et al., 1995). For some students, the predicament begs the question of whether a performance issue would exist if the particular test in use were not the standard to which students are held (Nelson, 1998; Raivetz, 1992). This is not to propose that an absence of standards and testing would be the answer; the point is to suggest that engagement and accountability ought not to be mutually exclusive objectives.

ASHA (2000) asserts that SLPs have curricular responsibilities on behalf of all students, especially students of minority group status or who are at risk for school failure. ASHA (2000, p. 1) proposes that SLPs can make a valuable contribution to the overall development of students in a school, district, or community by working with students with a range of abilities on "difficulties using language strategically to communicate, think, and learn." As the mandates of IDEA (1997) become fully implemented (ASHA, 1996, 1999, 2000; Mead, 1999), SLPs will be assessing and augmenting the abilities of students who are on their caseload as well as those who are not on caseload in order to meet curricular demands. But high-stakes, summative tests do not identify educationally handicapping conditions or suggest the academic supports that students at risk would need to receive in order to perform better on curriculum demands and on subsequent testing (Barkley, 2001). By and large, test outcomes do not point to useful guidelines for how to design ensuing instructional practices relative to future test expectations. Indeed, the clarity and usefulness of test score reports have been called into question (NEA, 2001b). This ambiguity may leave SLPs wondering how best to intervene to promote test achievement. Curriculum-based speech-language interventions that may prepare students for the demands of high-stakes testing are proposed.

INVOLVING SLPS IN INTERVENTIONS FOR STUDENTS WITH SPECIAL NEEDS WHO STRUGGLE WITH HIGH-STAKES TESTS

IDEA (1997) requires local and state education agencies to hold students with disabilities to the same standards as students without disabilities. Participation in assessment is an important aspect of equal access to education (NCEO, 2001). Special education and related services should be aligned with the general education curriculum, although curriculum modifications and instructional adaptations may be indicated for students with special needs (IDEA, 1997; NCEO, 2001). Standards-based reforms emphasize that every student, regardless of degree of disability, must work toward the expectations set for academic content (Silliman, 2000).

Failure on high-stakes testing may be caused by a lack of language-based information learned in and out of school (Popham, 1999). In some cases, test performance is related

to underlying oral language insufficiencies and hence poorer reading, writing, and school learning (Pershey, 2001). Identification of students in need and at risk should go beyond psychoeducational considerations and take high-stakes testing expectations into account (NCEO, 2001). This may entail new or expanded roles for SLPs as members of collaborative intervention teams whose focus is students' academic success. Intervention teams can establish a single set of shared functional and educationally relevant goals to be addressed by means of a collaborative, transdisciplinary approach to service delivery (Lyon & Lyon, 1980). In line with IDEA (1997) mandates, roles for SLPs include assessing and enhancing literacy-related and curriculum-relevant language abilities, encouraging the reciprocal relationship between spoken and written language with the goal of improving communication across all modalities, and "collaborating with others to shape or modify the curriculum for children with and without spoken or written language difficulties" (ASHA, 2000, p. 2).

Service within an academic context allows SLPs to consider not just the disorders, delays, differences, or sets of circumstances that place students at risk, but instead consider a whole child and the context in which the child must function all day (Preis & Schoenbrodt, 2000). SLPs' efforts might include providing consultation to teachers relative to curricular and instructional modifications, serving on regular education curriculum committees, assisting team members in preparing their communication-related goals and objectives for individualized education plans (IEPs), and using more test-relevant regular education materials when providing services (Pershey & Rapking, 2002). In these circumstances, SLPs and teachers need to define their roles and responsibilities, with, as Ehren (2000) proposes, SLPs being expert in language and knowledgeable about curriculum content and teachers being expert in curriculum content and knowledgeable about language. SLPs and teachers can mutually define how they will share the responsibility for helping students achieve language-dependent testing goals (NCEO, 2001).

SLPs and other team members should identify how successful test performance relates to three domains of proficiency that are commonly tested, namely, (a) knowledge of academic content (Kohn, 2000), (b) mastery of learning processes (Pottle, 2001), and (c) producing written products that are similar to those that the test will require or that support learning to write the types of products that the test will require (Pershey, 2001).

Academic content refers to the knowledge and information that students need to acquire. Rather than allowing testing to foster the memorization of facts (Kohn, 2000), teams might find that learning is deeper and more meaningful when academic content is presented in the context of interdisciplinary thematic units and lived-through learning experiences (Alvermann & Phelps, 1998; Taylor, 1996). For example, a unit on sea life can lead to an exploration of scientific concepts about marine animals and plants, pollution, and so forth, as well as a discussion of social studies concepts, such as the geography of the oceans and the economics of industries such as fishing and transportation. Literature selections might include biographies, nonfiction, and fiction. Music and art that relate to the sea theme could be experienced. SLPs can, for example, help students learn the vocabulary necessary to work through thematic units successfully.

Process mastery means that students are able to execute the learning skills or strategies that the team is teaching the students to use (Alvermann & Phelps, 1998; Farris, 1997; Gaskins, 1998; Harvey & Goudvis, 2000; Nelson, 1998; Pottle, 2001). SLPs can intervene to support the verbally mediated reasoning capabilities that are required to master learning skills and strategies. Here a skill means using a process to complete a task, but a strategy means knowing how to decide which skill to choose given a particular learning need (Meltzer, Roditi, & Stein, 1998). Examples of skills include how to use a Venn diagram, how to find a topic sentence in a paragraph, or how to write a prediction of how a story might end. An example of a strategy would be deciding if certain information would be displayed better on a pie chart or a bar graph.

Many high-stakes tests require students to produce a variety of written products. These products all have different pragmatic intents. SLPs can help students develop the macrostructural pragmatic language needed to compose a friendly letter, prepare a set of directions, write a summary or retelling of text, or write an expository piece that is descriptive or evaluative (Pershey, 1997, 2000). Students also have to use microstructural writing skills to apply vocabulary, grapheme–phoneme correspondences, morphosyntactic markers, and adequate sentence structure. SLPs can be a part of students' preparation for testing by assisting as students develop, for example, a personal dictionary for vocabulary related to a science unit or a chart that examines how root words change as affixes are added (Alvermann & Phelps, 1998; Farris, 1997; Harvey & Goudvis, 2000).

Each education agency requires that team members write IEP objectives that conform to the agency's policies. Examples of very generic IEP objectives that coincide with the need to prepare students to meet the academic content, skills processes, and written product demands established by test-regulated curricula are proposed. IEP objectives for academic content might read something like: "Student will participate in therapeutic intervention designed to reinforce classroom content with [improvement in] concept attainment shown 80% of the time." Process objectives might be worded: "Student will participate in activities designed to reinforce learning process skills introduced by the classroom teacher with 80% mastery of skill use." Or: "Student will select useful learning process strategies 80% of the time." Product objectives might state: "Student will participate in intervention designed to help student complete assigned products with completion of tasks on time 80% of the time."

To illustrate how SLPs can help students gain knowledge of academic content, master learning processes, and produce written products that reflect test demands, three case examples will be shared. The students' names are fictional.

Case Examples: Kate, Noah, and Max

Kate's case will be used to portray a student whose language therapy was focused on helping her master the academic content needed to participate in curriculum-based high-stakes assessment. Kate, in Grade 10, had receptive and expressive language weaknesses. In class, she had trouble defining words and using key words in sentences. Out of class, she sometimes did not comprehend language used in daily living, such as signs in stores, TV news stories, or package directions. Her IQ was roughly 85–90. Her mornings at school were spent in mainstream academic classes and her afternoons were spent in a prevocational program in food services. Her written language skills were in the Grade 4 to 6 range. Kate's SLP worked alone with her just before lunch time, 4 days per week for 20 minutes per session, on reviewing her notes from her morning classes. The SLP focused her interventions on curricular material that was likely to be tested in class and ultimately would coincide with her state's criterion-referenced test of state curriculum content. The purpose of intervention was for Kate to participate in reinforcement of classroom content in both oral and written formats. Kate and her SLP touched on concepts from all of her classes. They webbed concepts given in class; created flow charts to illustrate how one day's lesson proceeded into the next day's; kept a notebook of classroom vocabulary; and, outside of therapy time, kept a dialogue journal where they wrote back and forth to each other to express their understanding of classroom topics in their own words.

Noah's SLP worked with him to reinforce learning process skills introduced in the classroom. Noah was taught his academic subjects in a fifth-grade class for students with learning disabilities. Noah used limited syntax and vocabulary. He had poorly developed phonological skills that affected his speech and spelling. His academic skills were in the first- to second-grade range. During the 90 minutes per week that the SLP spent in Noah's academics classroom, Noah's learning disabilities teacher and the SLP worked together to teach skills processes that are often found on norm-referenced standardized testing. Noah worked to find the main idea in a short paragraph, retell the events of a short passage in their correct sequence, determine the characteristics of fictional versus nonfictional passages, and create titles for short passages.

Max's SLP worked with him on completing products that his teacher required. Max, in Grade 4, had difficulty completing classroom assignments because of problems in attention, concentration, language processing, and memory. He had trouble organizing his work and staying with tasks until their completion. He claimed to not understand the purposes of his assignments. As a classroom accommodation, Max was given fewer assignments and completed them with the assistance of the SLP. The SLP and Max's classroom teacher decided together which of the assignments given in class might be most useful for Max to do in preparation for tests of curriculum mastery. Working in a group with two other identified students from Max's class, Max and his SLP kept an assignment notebook, made written plans for executing various tasks, worked on assignments together, and wrote up a self-evaluation after completing assignments. Some of the assignments they finished included writing a book retelling, labeling a map of the continents and oceans of the world, writing a science lab report, and making a diagram of Max's family tree. They met just once a week for an hour so they would have a long enough period of time to get assignments done.

AN ADVOCACY ROLE: INVOLVING SLPS IN PROMOTING APPROPRIATE TESTING PRACTICES

SLPs who are involved with students and teachers in test preparation can become informed about prescriptions for appropriate, curriculum-relevant testing and about guidelines for testing special needs populations (NCEO, 2001). Accurate, repeated measurement of the content and processes that students have had the opportunity to learn and practice with the support of their SLPs is the primary goal (see, for example, "Texas Reading Initiative," 2002; "Texas's Reading Success Network: Year 1," 2002). The content and processes should reflect current educational reforms (U.S. Department of Education, 2002) that recognize discipline-based standards (e.g., language arts, mathematics).

For example, standards for curriculum content are central to the Goals 2000: Educate America Act of 1994 of the National Education Goals Panel (NEGP, 1994), as well as to the reauthorization of Title I (Linn, 1994) and to reforms proposed by the U.S. Department of Education (2002). The authors of these standards documents maintain that material that is tested should reflect curricular objectives that have importance instructionally, developmentally, intellectually, and culturally. The Goals 2000 panel (NEGP, 1994) also suggested standards for testing. Their recommendations specifically stated that (a) processes, concepts, and skills be tested, not curriculum content (e.g., facts in isolation); (b) performance standards reflect world-class standards, not minimum competency; and (c) students be given the opportunity to learn the concepts and skills and practice the performance tasks that will be included on the test. Even when tests are in line with these associations' recommendations, the issue remains whether (a) the test is testing educationally worthwhile objectives that ought to be the central part of the curriculum and (b) these objectives lend themselves to uncomplicated, direct, valid measurement. SLPs can be active members of teams that work to resolve these issues at various levels of authority: school, district, county, state, and so on.

Relative to students' needs in language and literacy development, joint standards are available through the National Council of Teachers of English (NCTE, 2001) and the National Association for the Education of Young Children (NAEYC, 1998) (both authored in conjunction with the IRA). It may be worthwhile for SLPs to consider ways in which tests correspond to these associations' guidelines for appropriate instruction, and then apply test demands and association guidelines to IEP objectives, collaboration plans, and test modification strategies (NCEO, 2001).

Students with special needs may take the same tests as their peers who are making typical academic progress, with or without accommodations, or they may take alternate assessments (IDEA, 1997). Accommodations may include changes in test presentation (e.g., read aloud, large print), response modality (oral as opposed to written response), setting, and timing or scheduling of test administration (NCEO, 2001). SLPs may be part of a team that devises and provides test accommodations or selects alternative assessments. High-stakes tests are often developed without educators' and SLPs' input; therefore, student assessment should be counterbalanced by using measures that reflect educators' and SLPs' involvement, such as portfolio assessment.

SLPs can address the practical need of demonstrating to teachers, parents, administrators, voters, politicians, and others that language proficiency is integral to meeting curricular demands and subsequently demonstrating competence on high-stakes testing (Pershey, 2001). Perhaps these efforts can help school speech-language therapy programs achieve full funding and full staffing (NEA, 2001a). Equitable access to enhanced educational opportunities and related services for students with academic language needs may be realized.

CONCLUSION

The NEA (2001b) contends that testing itself is not the route to improved learning, performance, and accountability. If students experience schooling as predominantly preparation for testing, they may not perceive education as a way to improve and enrich themselves and their communities. When standards of excellence are equated with test scores, other intellectual and ethical outcomes are potentially overlooked (Linn, 1994). The NEA proposes that education agencies evaluate multiple measures of school accountability in addition to testing outcomes, such as adequacy of resources and learning opportunities, absenteeism, dropout rate, parental involvement, and enrollment in advanced classes.

Supporting students with challenges and/or special needs as they undergo high-stakes testing is the responsibility of stakeholders at all levels—state, community, district, and school. Accountability can then reasonably be a shared effort by teachers, students, parents, special educators, and SLPs. Each has the capacity to help pupils deliver performance that is reflective of academic standards.

REFERENCES

Airasian, P. W. (1988). Symbolic validation: The case of state-mandated, high-stakes testing. *Educational Evaluation and Policy Analysis, 10,* 301–313.

Allington, R. L., & McGill-Franzen, A. (1992). Does high-stakes testing improve school effectiveness? *ERS Spectrum, 10,* 3–12.

Alvermann, D. E., & Phelps, S. F. (1998). *Content reading and literacy: Succeeding in today's diverse classrooms.* Boston: Allyn & Bacon.

American Educational Research Association. (2000). *AERA position statement concerning high-stakes testing in pre-K-12 education.* Retrieved December 15, 2000, from http://www.aera.org

American Speech-Language-Hearing Association. (1996). Inclusive practices for children and youths with communication disorders: Position statement and technical report. *Asha, 38*(Suppl. 16), 33–44.

American Speech-Language-Hearing Association. (1999). *Guidelines for the roles and responsibilities of the school-based speech-language pathologist.* Rockville, MD: Author.

American Speech-Language-Hearing Association. (2000). *Roles and responsibilities of speech-language pathologists related to reading and writing in children and adolescents.* Rockville, MD: Author.

Balow, I. H., Farr, R. C., & Hogan, T. P. (1992). *Metropolitan Achievement Test.* San Antonio, TX: Harcourt Brace Educational Measurement.

Barkley, R. (2001). The OEA [Ohio Education Association] on testing. *Ohio Schools, 79,* 12.

Barton, P. E. (1999). *Too much testing of the wrong kind, too little of the right kind in K–12 education: A policy information perspective.* Princeton, NJ: Educational Testing Service Policy Information Center.

Barton, P. E., & Coley, R. J. (1994). *Testing in America's schools: Policy information report.* Princeton, NJ: Educational Testing Service Policy Information Center.

Bobbett, G. C. (1993, April). *The impact of community/school characteristics on student outcomes: An analysis of report cards on schools.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Bond, L. A. (1995). *Norm-referenced testing and criterion-referenced testing: The differences in purpose, content, and interpretation of results.* Washington, DC: Office of Educational Research and Improvement.

Commission on Behavioral and Social Sciences and Education. (1998). *Equivalency and linkage of educational tests interim report.* Washington, DC: National Academy Press.

CTB/McGraw-Hill. (1992). *California Achievement Test.* Monterey, CA: Author.

CTB/McGraw-Hill. (1998). *TerraNova.* Monterey, CA: Author.

Denoyer, R., & White, M. (1992). *Proficiency testing and the income gap.* Washington, DC: U.S. Department of Education. (ERIC Document Reproduction Service No. ED35064)

Ehren, B. J. (2000). Maintaining a therapeutic focus and sharing responsibility for student success: Keys to in-classroom speech-language services. *Language, Speech, and Hearing Services in Schools, 31,* 219–229.

Etsey, Y. K. (1997, March). *Teachers' and school administrators' perspectives and use of standardized achievement tests: A review of published research.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Falk, B. (2000). *The heart of the matter: Using standards and assessments to learn.* Portsmouth, NH: Heinemann.

Farris, P. J. (1997). *Language arts: Process, product, and assessment.* Madison, WI: Brown & Benchmark.

Gallagher, M. P. (1993, April). *Proficiency testing and poverty: Looking within a large urban district.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Gaskins, I. W. (1998). There's more to teaching at-risk and delayed readers than good reading instruction. *The Reading Teacher, 51,* 534–547.

Glaser, R. (1994). Criterion-referenced tests: Part II. Unfinished business. *Educational Measurement: Issues and Practice, 13,* 27–30.

Gordon, S. P., & Reese, M. (1997). High-stakes testing: Worth the price? *Journal of School Leadership, 7,* 345–368.

Haladyna, T., Haas, N., & Allison, J. (1998). Continuing tensions in standardized testing. *Childhood Education, 74,* 262–273.

Harvey, S., & Goudvis, A. (2000). *Strategies that work: Teaching comprehension to enhance understanding.* York, ME: Stenhouse.

Heubert, P. P., & Hauser, R. M. (1999). *High stakes: Testing for tracking, promotion, and graduation.* Washington, DC: National Academy Press.

High stakes testing and social promotion, 106th Cong., 1 (1999) (testimony of Robert M. Hauser).

Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1996). *Iowa Tests of Basic Skills.* Itasca, IL: Riverside.

Individuals with Disabilities Education Act Amendments of 1997, Pub. L. No. 1015-17, 20 U.S.C. 1400 et seq.

International Reading Association. (1999). High stakes assessments in reading. *The Reading Teacher, 53,* 257–263.

Jones, L. V. (1997). *National tests and education reform: Are they compatible?* Washington, DC: Office of Policy and Planning.

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research, 64,* 425–461.

Kohn, A. (2000). *The case against standardized testing.* Portsmouth, NH: Heinemann.

Koretz, D. M. (1991, April). *The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests.* Paper presented at the annual meetings of the American Educational Research Association, Chicago, IL and The National Council on Measurement in Education, Chicago, IL.

Ladson-Billings, G. (1994). *The dreamkeepers.* San Francisco: Jossey-Bass.

Lanese, J. F. (1992, April). *Statewide proficiency testing: Establishing standards or barriers?* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Linn, R. L. (1994). *Assessment-based reform: Challenges in educational measurement.* Princeton, NJ: Educational Testing Service.

Lyon, S., & Lyon, G. (1980). Team functioning and staff development: A role release approach to providing integrated educational services for severely handicapped students. *JASH (Journal of the Association for Persons with Severe Handicaps), 5,* 250–263.

Maddaus, G. F. (1988). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education, 65,* 29–46.

Manning, M. L., Lucking, R., & MacDonald, R. H. (1995). What works in urban middle schools. *Childhood Education, 71,* 221–224.

McGee, G. W. (1997, March). *What state tests test.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Mead, J. F. (1999). The reauthorization process of the Individuals with Disabilities Education Act: Expressions of equity. *Journal for a Just and Caring Education, 5,* 476–492.

Meisels, S. (1989). *Testing, tracking, and retaining young children: An analysis of research and social policy.* Washington, DC: National Center for Education Statistics.

Meltzer, L., Roditi, B., & Stein, J. (1998). Strategy instruction: The heartbeat of successful inclusion. *Perspectives, 24,* 10–13.

National Assessment of Educational Progress. (1998). *Long-term trends in student reading performance.* Available: http://nces.ed.gov/pubs98/98464.html

National Assessment of Educational Progress. (1999a). *NAEP 1998 reading report card for the nation and the states.* Washington, DC: US Department of Education Office of Educational Research and Improvement. Retrieved December 15, 2000, from nces.ed.gov/nationsreportcard/pubs/main1998/1999500.html

National Assessment of Educational Progress. (1999b). *Trends in academic progress.* Washington, DC: US Department of Education Office of Educational Research and Improvement.

National Assessment of Educational Progress. (2000). *The NAEP guide.* Retrieved December 15, 2000, from http://nces.ed.gov/nationsreportcard/guide/ques2.shtml#2000reading

National Association for the Education of Young Children. (1998). *Overview of learning to read and write: Developmentally appropriate practices for young children.* Retrieved July 1, 2001, from http://www.naeyc.org/resources/position_statements/psread0.htm

National Center on Educational Outcomes. (2001). *Special topic areas.* Retrieved July 1, 2001, from http://www.coled.umn.edu/nceo

National Council of Teachers of English. (2001). *Standards for the English language arts.* Retrieved July 1, 2001, from http://www.ncte.org/standards

National Educational Research Policy and Priorities Board. (1999). *Investing in learning: A policy statement on research in education.* Washington, DC: US Department of Education.

National Education Association. (2001a). *Legislative action center.* Retrieved July 1, 2001, from http://www.nea.org/lac

National Education Association. (2001b). *Testing plus.* Retrieved July 1, 2001, from http://www.nea.org/testingplus

National Education Goals Panel. (1994). *Goals 2000 The national education goals report: Building a nation of learners 1994.* Washington, DC: Author.

Nelson, N. W. (1998). *Childhood language disorders in context.* Boston: Allyn & Bacon.

Pershey, M. G. (1997). Teaching pragmatic language awareness as an integral aspect of reading and language arts instruction. *Reading Horizons, 37*(4), 299–314.

Pershey, M. G. (2000). Teaching communicative competence and metalinguistic awareness during reading/language arts instruction. *Primer, Journal of the Massachusetts Reading Association, 28*(3), 1–11.

Pershey, M. G. (2001, November). *State mandated literacy tests: Exigency for African American students.* Poster presented at the annual convention of the American Speech-Language-Hearing Association, New Orleans, LA.

Pershey, M. G., & Rapking, C. (2002). Collaborative speech-language services in urban schools. *Academic Exchange, 6*(3), 75–80.

Phelps, R. P. (1999). Why testing experts hate testing. *Fordham Report, 3,* 1–41.

Popham, W. (1994). The instructional consequences of criterion-referenced clarity. *Educational Measurement: Issues and Practice, 13,* 15–18, 30.

Popham, W. (1999). Why standardized tests don't measure educational quality. *Educational Leadership, 56,* 8–15.

Pottle, P. (2001). Coping with a process-oriented mandate. *Primary Voices K-6, 9,* 32–37.

Preis, J., & Schoenbrodt, L. (2000, November). *Service learning: A method to develop well-rounded professionals.* Poster presented at the annual convention of the American Speech-Language-Hearing Association, Washington, DC.

Raivetz, M. J. (1992). Can school districts survive the politics of state testing initiatives. *NASSP Bulletin, 76,* 57–65.

Robinson, M. A., & Moore, M. H. (1992, April). *State proficiency tests and standardized tests.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Roderick, M. (1995). *Grade retention and school dropout: Policy debate and research questions.* Phi Delta Kappa Center for Evaluation, Development, and Research, Research Bulletin No. 15. Retrieved December 15, 2000, from http://www.pdkintl.org/edres/resbul15.htm

Shepard, L. A., & Dougherty, K. C. (1991, April). *Effects of high-stakes testing on instruction.* Paper presented at the Annual Meetings of the American Educational Research Association, Chicago, IL, April 3–7, 1991 and The National Council on Measurement in Education, Chicago, IL.

Sheridan, C. (2000, August 27). Governor's panel focuses on—gasp—teachers. *The Plain Dealer,* pp. 1–2G.

Silliman, E. R. (2000). From the editor. *Language, Speech, and Hearing Services in Schools, 31,* 211.

Stanford Achievement Test. (1996). San Antonio, TX: Harcourt Brace Educational Measurement.

Stephens, S. (2000, December 10). Kindergarten shifts gears, trading blocks for books. *The Plain Dealer,* p. 1, 5B.

Stotsky, S. (1997). State English standards. *Fordham Report, 1*(1), Appendix D. Retrieved December 15, 2000, from www.edexcellence.net/stotsky/stotd3.html

Stroud, R. E. (1995). *Correlations between the Metropolitan Achievement Tests, Seventh Edition, and the Ohio Ninth-Grade Proficiency Tests.* Washington, DC: U.S. Department of Education. (ERIC Document Reproduction Service No. ED394992)

Tapper, R. (1997). *The problem of high stakes assessment in public education.* Newark, DE: International Reading Association.

Taylor, T. R. (1996). *Reshaping the curriculum: Using an integrated, interdisciplinary approach.* Oak Brook, IL: Curriculum Design for Excellence.

Texas Reading Initiative. (2002). Retrieved July 1, 2001, from http://www.starcenter.org/ppt/rsn/sld008.htm

Texas's Reading Success Network: Year 1. (2002). Retrieved July 1, 2001, from http://www.starcenter.org/ppt/rsn/sld008.htm

United States Department of Education. (2002). *FY 2003 budget summary. Elementary and secondary education. State assessments and enhanced assessment instruments.* Retrieved August 1, 2002, from http://www.ed.gov/offices/OUS/Budget03/Summary/SectionII/A.html#2

Webb, M. W. (1995, April). *Policy considerations in developing standards and assessments for large, diverse school districts.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Westbury, M. (1994). The effect of elementary grade retention on subsequent school achievement and ability. *Canadian Journal of Education, 19,* 241–250.

Winfield, L. F. (1990). School competency testing reforms and student achievement: Exploring a national perspective. *Educational Evaluation and Policy Analysis, 12,* 157–173.

Contact author: Monica Gordon Pershey, Department of Speech and Hearing, Cleveland State University, 2121 Euclid Avenue, Cleveland, OH 44115-2214. E-mail: m.pershey@csuohio.edu