

Alternate Assessment: Have We Learned Anything New?

ELIZABETH TOWLES-REEVES

HAROLD KLEINERT

MONICAH MUHOMBA

University of Kentucky

ABSTRACT: *Since the first review of alternate assessment literature by Browder, Spooner, Algozzine, et al. (2003), educational research related to alternate assessment has rapidly expanded as state and federal policies have required the inclusion of students with disabilities, including students with the most significant cognitive disabilities, in state and school accountability indexes. The purpose of this literature review was to examine the empirical studies conducted since the conception of alternate assessment to evaluate the progress made in this area, using a research framework proposed by Browder et al. We discuss future research directions, particularly as these relate to recent changes in federal policy under No Child Left Behind as well as implications for teachers and practitioners.*

Large-scale educational assessment has become an increasingly important strategy over the past decade for holding schools accountable for the teaching and learning of all students. One element of large-scale educational assessment within each state is that state's alternate assessment system. As a result of the Individuals With Disabilities Education Act of 1997 (IDEA), all students are required to be included in state and district educational assessments [612(a)(17)(A)]. For those students who cannot participate in the general assessment system, even with appropriate accommodations, an alternate assessment is required [612(a)(17)(A)(i-ii)]. IDEA 1997 allows for the development of alternate assessments within states to ensure students with the most significant cog-

nitive disabilities are included in state and district assessments in a way that is fair and equitable; these requirements are reiterated in the Individuals With Disabilities Education Improvement Act of 2004 (IDEA, 2004) and the No Child Left Behind Act of 2001 (NCLB, 2001).

In 2003, Browder, Spooner, Algozzine, et al. published a review of empirical studies related to alternate assessments based on alternate achievement standards (AA-AAS). This review included 19 databased articles published in peer-refereed journals or in press through December of 2002. Their review provided a basis for illustrating what educational researchers knew and needed to know about alternate assessments in relation to measuring the progress of students (i.e., achievement) with the most significant cognitive disabilities through large-scale educational assessment systems.

Since the publication of the Browder, Spooner, Algozzine, et al. (2003) article, the field of educational research related to alternate assessment has rapidly evolved as state and federal policies have required the inclusion of students with disabilities, including students with the most significant cognitive disabilities, in state and school accountability indexes. When Browder and colleagues (2003) published their literature review, only 19 empirical studies had been conducted since the conception of alternate assessment in the early 1990s. Since then, 23 additional empirical studies have been published. Given that the number of databased studies on AA-AAS has more than doubled since December of 2002, and with recent changes in federal policy in respect to alternate assessment, another review of the literature in AA-AAS was timely.

DEFINING AA-AAS

Because of the changing landscape of alternate assessment, it is important to highlight what currently defines AA-AAS. AA-AAS is the primary method through which students with the most significant cognitive disabilities participate in measures of educational assessment and school accountability (Quenemoen, Rigney, & Thurlow, 2002). Roeber (2002) outlined three widely used assessment approaches:

- A portfolio, or body of evidence, approach is a purposeful and systematic collection of student work evaluated and judged against predetermined scoring criteria.
- A checklist approach requires teachers to identify whether students are able to perform certain skills, tasks, or activities. In this approach, scores are based on the number of skills the student is able to perform successfully.
- A performance assessment (or performance event) approach is a direct measure of a skill in a one-to-one assessment format (e.g., the student responding to questions about plot in a preselected, grade-level, fictional text).

Although most state departments of education have chosen to use one of the three alternate assessment approaches as outlined by Roeber

(2002), some have developed hybrid approaches merging two of the three approaches. For example, a state may use a checklist approach and also require a body of evidence to supplement the checklist. State departments of education also have the freedom to build their alternate assessment system to fit appropriately within the state's own large-scale educational assessment system. Consequently, differences in the approaches across states (e.g., portfolio vs. performance event vs. checklist) and differences within approaches (e.g., variations in portfolio assessments from one state to the next) are typical.

POPULATION OF LEARNERS FOR WHOM AA-AAS ARE DESIGNED

As with any assessment, especially one designed for a subset of the population of learners with disabilities, it is critical for validity purposes to understand for whom the assessment is designed and who actually participates in that assessment. AA-AAS are reserved for a small percentage of the student population (students with the most significant cognitive disabilities) for whom traditional paper and pencil assessments, even with appropriate accommodations, would be an inappropriate measure of student progress within the general education curriculum. For those students who cannot participate in regular assessments, even with accommodations, states (or districts, in the case of local assessments) must develop alternate assessments. "In general, the Department [of Education] estimates that about 9 percent of students with disabilities (approximately one percent of the total student population) have significant cognitive disabilities that qualify them to participate in an assessment based on alternate achievement standards" (U.S. Department of Education, 2005a, p. 23). Researchers have noted that students taking this assessment typically have special education labels such as autism, mental retardation, and/or multiple disabilities (U.S. Department of Education, 2003). However, not all students with these labels will require an alternate assessment, and students with other special education labels may also qualify for an AA-AAS.

ALTERNATE ACHIEVEMENT STANDARDS

Alternate achievement standards are "an expectation of performance that differs in complexity from a grade-level achievement standard" (U.S. Department of Education, 2005a, p. 20). Alternate achievement standards must be defined to meet four conditions: (a) must be aligned with the state's academic content standards; (b) must describe at least three levels of achievement (i.e., basic, proficient, and advanced); (c) must include descriptions of competencies associated with each level of achievement; and (d) must include assessment or cut scores that differentiate between achievement levels. A description of the rationale and procedures used to determine the levels of achievement must also be included (U.S. Department of Education, 2005a). Further, though alternate achievement standards may differ in complexity from grade-level achievement standards, they must still be linked to grade-level content. In its NCLB Peer Review Guidance for states, the U.S. Department of Education (2004) has made this linkage to grade-level context explicit:

For alternate assessments in grades 3 through 8 based on alternate achievement standards, the assessment materials should show a clear link to the content standards for the grade in which the student is enrolled although the grade-level content may be reduced in complexity or modified to reflect pre-requisite skills. (p. 15)

This linkage to grade-level content standards, even for AA-AAS, represents a new requirement for alternate assessments, one that was not in place at the time of the Browder, Spooner, Algozzine, et al. (2003) review and which will undoubtedly shape future research in alternate assessment.

We should also note that federal law permits states to develop two other types of alternate assessments: alternate assessments on grade-level achievement standards (or AA-GLAS) and alternate assessments on modified achievement standards (or AA-MAS). Students participating in the AA-GLAS have disabilities that are typically not cognitive in nature but require an alternate format or context to demonstrate their knowledge. The AA-GLAS is, in general, not appropriate for students with the most significant cognitive dis-

abilities, as performance on that assessment is judged against the same grade-level achievement standards as students in the regular assessment (Wiener, 2006). Similarly, the AA-MAS is also not appropriate for students with the most significant cognitive disabilities, as this assessment is intended to serve students whose progress toward proficiency would best be measured against modified achievement standards. Nonregulatory guidance states an AA-MAS is appropriate for "a student whose disability has precluded the student from achieving proficiency, as demonstrated by objective evidence of the student's performance and whose progress is such that, even if significant growth occurs, the student's IEP team is reasonably certain that the student will not achieve grade-level proficiency within the year covered by the IEP" (U.S. Department of Education, 2007, p. 53). Most often, students with the most significant cognitive disabilities participate in AA-AAS (U.S. Department of Education, 2005a) as this option allows for reduced depth, breadth, and complexity in the alternate academic achievement standards by which a student's progress is measured.

This review will focus exclusively on empirical research regarding AA-AAS for the following reasons: (a) Browder, Spooner, Algozzine, et al.'s (2003) study entirely addressed AA-AAS; (b) all states have an AA-AAS, whereas only a limited number of states have an AA-GLAS (Thompson, Johnstone, Thurlow, & Altman, 2005); (c) AA-MAS systems are now just emerging; and (d) all of the peer-reviewed literature to date has focused on AA-AAS.

REVIEW OF PAST LITERATURE

Browder, Spooner, Algozzine, et al. (2003) reported that at the time of their publication, there were insufficient data "to report with confidence that alternate assessment will live up to its promises" (p. 51) of (a) greater consideration of students with disabilities in school and state policy decisions, (b) overall increased expectations for individuals with disabilities, (c) increased access to the general curriculum and state academic content standards for students with disabilities, and (d) improvement of instructional decision

making at both the teacher and classroom level for students with disabilities. As a result, Browder and colleagues suggested six categories or themes for future research related to alternate assessment:

- Validate performance indicators with content area experts and stakeholders.
- Use a format for alternate assessment that produces data for instructional decisions.
- Link alternate assessments to the individualized education program (IEP) so students and parents can participate in setting the level of expectation.
- Train teachers in how to incorporate alternate assessment in daily practice.
- Use best measurement practice for scoring and reporting alternate assessments and collecting and reporting data on technical quality.
- Use alternate assessment outcomes for program evaluation and ongoing quality enhancement.

Browder, Spooner, Algozzine, et al. (2003) recommended that research in these six areas would enable the field to determine if alternate assessment had attained its promise. The purpose of the current review was to integrate all the literature conducted since the conception of AA-AAS (including those articles reviewed in the original Browder et al. review) to determine how well the current literature on AA-AAS addresses these six categories. This review will also pose additional research questions as well as examine the significance of future research that may evolve as a result of changes in federal policy.

METHODS

The authors began by including the appropriate articles in the original Browder, Spooner, Algozzine, et al. (2003) literature review that were published after the IDEA (1997) final regulations were propagated (total of 17 articles). To identify additional empirical literature appropriate for inclusion in this review, the authors used the following electronic search engines: Exceptional Child EdRes, EBSCOhost, ERIC, and PsycINFO using the descriptors of alternate assessment, large-scale assessment, and students with the most significant cognitive disabilities. In addition, we searched the reference lists of the articles appro-

priate for inclusion in this review. At the same time, leading experts conducting research in alternate assessment were contacted and asked for summaries of their databased research that had been accepted for publication in peer-refereed journals. Criteria for inclusion in this review were as follows:

- The article had to have at least one measure directly related to AA-AAS.
- The research had to use a quantitative or qualitative research design or provide program evaluation data.
- The research was published or in press in a peer-reviewed journal or part of the knowledge base developed by the National Center on Educational Outcomes (NCEO) prior to July 2007.

After applying these criteria to all the studies found in the literature search and including the 17 articles from the original Browder, Spooner, Algozzine, et al. (2003) review, a total of 40 databased studies were found that were appropriate for this review.

HAVE RECOMMENDATIONS BEEN MET?

The following six sections outline the progress that has been made toward understanding the research related to AA-AAS as recommended by Browder, Spooner, Algozzine, et al. (2003). After each section, we summarize the status of the field given what we know from the existing literature and policy guidance.

Recommendation 1: Validate Performance Indicators With Content Area Experts and Stakeholders. Ten articles have been published that investigate validating performance indicators with content area experts and stakeholders. Kleinert and Kearns (1999) first validated the performance indicators and learner outcomes of Kentucky's alternate assessment with 44 national authorities and experts in best practices for students with moderate and severe cognitive disabilities. A high degree of congruence was found between best practices and Kentucky's performance indicators, but experts expressed concerns about low expectations for students with significant cognitive disabilities and alignment of Kentucky's alternate assessment with grade-level academic content

standards important for all students to learn. Five years ago, this was the only research that had attempted to validate performance indicators. With the passage of NCLB (2001), the focus on teaching students with the most significant cognitive disabilities grade-level academic content, and the assessment of students on this content, the importance of alignment (or linkage) of alternate academic achievement standards with grade-level content standards has become a critical issue.

Since the first study by Kleinert and Kearns (1999), other researchers have begun to study the current status of alignment for alternate assessments across the nation. Browder et al. (2004) examined the alignment of performance indicators in 31 states to national standards and curricula. Results suggested a strong focus on academic skills (i.e., reading and math) for alternate assessments with a more complementary focus on functional skills. In addition, Browder, Spooner, Ahlgrim-Dezell, et al. (2003) investigated curricular philosophies and alignment of AA-AAS with states' performance indicators. In this study, states demonstrated a blend of academic and functional skills in creating their performance indicators. Most important, in states where academic philosophies (i.e., instruction on grade-level academic content) were stressed more than functional philosophies (i.e., instruction on life skills), clear links to the academic content standards were evident. Building on these studies, Browder, Ahlgrim-Dezell, et al. (2005) conducted an analysis of alternate assessments and their alignment with the state's academic content standards. Findings indicated most states measured academic domains such as reading and math on their alternate assessment but a few states only measured functional skills.

Johnson and Arnold (2004) conducted a validity study of Washington State's alternate assessment. For portfolios submitted in the 2001 to 2002 school year, these researchers found that "more than 75% of all portfolio entries were in some way connected to the state standards, with the highest percentage (81.4%) in reading and the lowest in math (75.1%)" (p. 270). However, they noted that these percentages may well have been inflated in that they represented the teacher's interpretation of the relationship of the IEP skill to the content standard, and as such, these appar-

ent connections may not have represented a well-defined alignment of portfolio content to state standards. In a follow-up study, Johnson and Arnold (2007) again found that although a very high percentage of the alternate assessment portfolios indicated a connection to general state standards (ranging from 88% in seventh-grade math to 97% in eighth-grade science), there was tremendous variation in how the teacher-targeted assessment skills evidenced state standards. For example, in fourth-grade writing, targeted skills ranged from "writing an eight-sentence paragraph on a given topic" to "participate in art by coloring a 4-in. shape." Johnson and Arnold suggested that for many portfolios, the connection to state standards was a superficial one.

Kohl, McLaughlin, and Nagle (2006), in a study of 16 randomly selected states, found that 14 of these states held students with significant cognitive disabilities to the same general content standards as all students, though states varied considerably in how these standards were operationalized for students with significant cognitive disabilities to reflect grade-level content or specific performance descriptors or indicators. According to Kohl et al., "In turn, these content adjustments have produced simplified standards that consist primarily of functional academic skills Furthermore, the adjusted standards do not appear to follow a logical sequenced curriculum across the grade levels" (p. 120). It is interesting to note that even though 14 of the 16 states in this study reported using a stakeholder group to adjust the content standards for students with significant cognitive disabilities, the only constant members of these groups across each of these states were special education teachers and administrators. General education teachers, curriculum specialists, and parents were involved to a lesser extent.

Roach, Elliott, and Webb (2005) investigated the alignment of Wisconsin's alternate assessment with state academic content standards through the use of Webb's (2002) criteria for alignment: (a) categorical concurrence (the presence of at least six assessment items for each content standard), (b) range of knowledge (more than half of the objectives from any given standard are assessed), (c) balance of representation of assessment items across instructional objectives, and (d) depth of knowledge (at least 50% of the assess-

ment items must reflect a depth or application of knowledge at least at the level of the actual objectives). Although that state's alternate assessment did not evidence strong categorical concurrence in language arts and science, range of knowledge was acceptable across each of the subject areas for the alternate assessment, as was the alternate assessment's balance of representation across items. Although depth of knowledge might be expected to be lower for alternate assessments than for regular assessments (because achievement standards can be reduced in complexity for alternate assessments), Roach et al. found acceptable depth of knowledge ratings for 13 of the 25 standards assessed across reading, math, science, and social studies for the Wisconsin alternate assessment.

It is interesting to note . . . the only constant members of these groups across each of these states were special education teachers and administrators.

Using an earlier version of Webb's (1997) alignment framework, Flowers, Browder, and Ahlgrim-DeLzell (2006) also took an in-depth look at the match between three states' alternate assessments and academic content standards. Findings revealed these assessments are aligned but only capture a narrow range and depth of the content standards, possibly suggesting lowered academic expectations for students. However, it should be noted that this study was conducted before states were required to make explicit links to grade-level content standards in their AA-AAS, and this requirement may result in different findings in future alignment studies.

Finally, Karvonen, Wakeman, Flowers, and Browder (2007) investigated the development and pilot of a curriculum survey for use in conducting alignment studies and for teacher professional development. The purpose of this research was to investigate instructional alignment (the type, number, and complexity of academic skills being taught to students with the most significant cognitive disabilities). Preliminary results suggest this is a useful tool for conducting alignment studies and as a self-assessment for teachers. Although future work is necessary to determine the

value of the instrument for each of these purposes, the use of a curriculum survey tool as an essential element within alignment studies takes the notion of alignment from a simple correspondence of state standards to test content to the broader context of what is actually taught at the classroom level.

Together, these 10 studies characterize the status of the field related to validating the performance indicators of alternate assessments. These studies suggest the need for experts in both significant cognitive disabilities and academic content domains to establish alignment criteria, as well as the necessary amount of alignment, for students participating in AA-AAS. These studies further suggest that other key stakeholders, including assessment experts, parents, and school administrators, are not always included in this process. As a group, these studies also suggest that alignment to state standards for students with significant cognitive disabilities, while present, is highly variable in how it is operationalized within individual states, is often based on the judgment of the teacher, and at times, represents a distant connection at best.

Criteria for alignment should facilitate access to the general curriculum and reflect high expectations for students with the most significant cognitive disabilities. Along these lines, Browder et al. (2007) have recently proposed a conceptual definition of alignment and explicated seven total criteria for consideration when linking instruction and assessment to grade-level academic content standards. Their intended purpose was to "create a conceptual framework for building consensus about general curriculum access and conducting future research" (p. 12) in the area of alignment for AA-AAS.

Furthermore, instrument and process development will be important in helping states conduct alignment studies and reflect on teacher professional development for understanding and creating assessments aligned (or linked) to grade-level academic content standards for students with the most significant cognitive disabilities. Emerging work (Karvonen et al., 2007) will be useful for informing instrument development not only for alignment but for other purposes such as teacher self-assessments. The Links for Academic Learning model developed by Flowers, Wakeman, Browder, and Karvonen (2007) is another re-

cently developed model for conducting alignment studies with AA-AAS that may be of assistance to states and districts in this process.

Recommendation 2: Use a Format for Alternate Assessment That Produces Data for Instructional Decisions. Leading researchers in alternate assessment (Browder, Spooner, Algozzine, et al., 2003; Kleinert & Kearns, 2004) have noted that alternate assessment should yield student performance data that can guide and improve daily instruction. To date, no published research has directly investigated the use of alternate assessment to directly inform instructional decisions. Although the field of special education has produced a strong research base on the value of ongoing, data-based decisions (see Farlow & Snell, 1994), this research has not focused on academic targets for students with significant cognitive disabilities. Only recently have scholars begun to understand applications of academic content to students with significant cognitive disabilities (Browder, Spooner, Ahlgrim-Dezell, Harris, & Wakeman, 2008; Browder, Wakeman, Spooner, Ahlgrim-Dezell, & Algozzine, 2006; Courtade, Spooner, & Browder, 2007). Research is clearly needed to determine how alternate assessment can guide instructional decisions on academic content for students with the most significant cognitive disabilities (Perner, 2007).

Recommendation 3: Link Alternate Assessment to the IEP so Students and Parents Can Participate in Setting the Level of Expectation. Although it is important for IEPs to reflect standards-based skills (U.S. Department of Education, 2004, 2007) to improve access to the general curriculum for students with significant cognitive disabilities, the five studies conducted in this area to date have revealed a lack of a clear link between alternate assessment and the IEP. Turner, Baldwin, Kleinert, and Kearns (2000) first investigated the extent to which scores on Kentucky's alternate assessment were correlated with IEP quality. Results indicated these two variables were not significantly correlated, suggesting teachers did not see the relationship between the alternate assessment and IEPs. Similarly, White, Garrett, Kearns, and Grisham-Brown (2003) found no relationship between "assessment outcomes and the quality of a student's IEP or overall instructional programming" (p. 205), even though students who had

greater opportunities for developing communication and social skills had better outcomes on the alternate assessment. Thompson, Thurlow, Esler, and Whetstone (2001) examined IEP forms from 41 states to determine if these forms included academic content standards or general curriculum expectations. Results suggested very few states even mentioned state standards on their IEP forms, whereas only 30 of 41 states had forms that included choices for assessment options, a mandated part of the IEP process.

Towles-Reeves, Garrett, Burdette, and Burdige (2006) found that alternate assessments did influence the development of students' IEPs, but the influence of alternate assessments on IEP development was significantly less than the influence of the assessment on teachers' daily instructional practices. Finally, Towles-Reeves and Kleinert (2006) found that the majority of teachers (58.6%) in one state's alternate assessment found no overall influence of the alternate assessment on IEP development; however, considerably more teachers (34.5%) in that study reported a positive influence of the alternate assessment on IEP development than those who reported a negative influence (6.1%).

As a whole, these five studies clearly indicate additional research is needed to examine the link between IEPs and alternate assessment, especially because new guidance requires IEPs to include grade-level academic content standards for AA-MAS (U.S. Department of Education, 2007) and for increased access to the general curriculum for students with the most significant cognitive disabilities who typically take AA-AAS (U.S. Department of Education, 2005a). Although the content of IEPs for students with significant cognitive disabilities will necessarily be broader than the content of the AA-AAS (because the IEP must also address the student's other educational needs that arise from the student's disability and the AA-AAS is based on the state content standards for all students, as mandated by NCLB, 2001), there appears to be a need to train IEP teams to develop standards-based IEPs (Thompson et al., 2001) that are linked to grade-level content standards. As a result, teachers and parents can grasp how academic content standards can be appropriately individualized even for students with the most significant cognitive disabilities.

Recommendation 4: Train Teachers in How to Incorporate Alternate Assessment in Daily Practice.

Research in alternate assessments has demonstrated the need to train teachers in how to incorporate alternate assessment into their daily instruction to ensure that students with significant cognitive disabilities access the general curriculum in meaningful ways. Destefano, Shriner, and Lloyd (2001) studied the effectiveness of training on teachers' knowledge of participation and accommodation decisions related to large-scale assessments for students with disabilities. Findings suggested that after training, teachers expressed confidence in making accommodation decisions and also indicated an improved understanding of the relationship between participation/accommodation, curriculum, and instruction. Kampf, Horvath, Kleinert, and Kearns (2001) investigated the relationship of time and effort to scores on alternate assessments. Results indicated instructional variables (i.e., extent to which portfolio items were embedded into daily instruction and extent to which students were involved in their own portfolio assessments) were significantly related to students' scores. Significantly less related to students' scores, however, was the amount of time spent on actually completing the assessment. These authors concluded that encouraging teachers to incorporate the assessment into daily instruction is important to improving scores on alternate assessments.

Results from a survey of teachers' perceptions of alternate assessments in five states revealed that increase in paperwork and demand on time was the most significant impact of alternate assessment (Flowers, Ahlgrim-Delzell, Browder, & Spooner, 2005). These authors noted that only 28% of their respondents reported that their students had greater access to the general curriculum, only 25% reported increased progress on IEP objectives, and only 25% reported an overall better quality of education as a result of the alternate assessment in their respective states, suggesting that alternate assessments have not resulted, at least in the view of these teachers, in significant changes in instructional practices. Yet, results from the same survey also revealed that when teachers perceived the alternate assessment to contribute to school accountability, more teachers reported a positive impact. The authors suggested

that reducing the amount of paperwork and providing teachers with more models of how to address state standards in appropriate ways could be helpful in achieving the purpose of alternate assessment. For alternate assessment to improve the quality of education for students with severe disabilities, it is important that teachers perceive that students benefit from the use of alternate assessment procedures and instruments.

This need for teacher training was also echoed by Horvath, Kampf-Bohach, and Kearns (2005). These researchers described the degree to which accommodations used by students with deaf-blindness were documented in the IEP and also implemented with fidelity during both instruction and assessment. Results from this study indicated that most students with deaf-blindness used different accommodations in the classroom than in assessment, even though these students used a wide variety of accommodations both in assessment and in daily classroom instruction. Horvath et al. also noted significant discrepancies between the accommodations that were listed on the IEP and those actually put into practice, suggesting a need for educator training in the use of accommodations during instruction, as well the identification of accommodations on the IEP. The authors concluded that it is crucial for researchers and educators to ensure that students with deaf-blindness are provided equal access to both the curriculum and the test content.

Browder, Karvonen, Davis, Fallin, and Courtade-Little (2005) found that when teachers received training on instructional practices and data-based decision making, students' alternate assessment scores improved in comparison to students of teachers who did not receive such training. The students of the teachers who received the training also made greater progress on their IEP objectives. Findings from this study provided the "first evidence that alternate assessment scores can be improved through training teachers in instructional variables" (Browder, Karvonen, et al., p. 277). The authors suggested considerable training will be necessary in the years to come if teachers are to understand how to teach and assess students with the most significant cognitive disabilities in language arts, math, and science.

In a case study of seven teachers and students with a history of positive scores on one state's al-

ternate assessment, Karvonen, Flowers, Browder, Wakeman, and Algozzine (2006) found that teachers who provided extensive amounts of direct instruction on targeted assessment skills to their students, took frequent and even daily data, ensured data accuracy, and enabled students to participate in monitoring their own learning through strategies such as self-evaluation had students who scored well on that alternate assessment (i.e., proficient or higher). This finding echoes that of Kampfer et al. (2001) that the assessment should be integrated within the context of daily instruction and students should have direct involvement in their own assessment.

Finally, in a statewide teacher survey in a rural, Midwestern state, Towles-Reeves and Kleinert (2006) found that 44% of responding teachers perceived that the AA-AAS did have a positive impact on daily instruction, whereas only 16% of teachers perceived a negative impact. However, 39% of the teachers indicated that the alternate assessment had no impact on daily instruction. For those teachers who did rate the alternate assessment as having a positive impact on daily instruction, the components or dimensions of the alternate assessment rated as having the strongest impact on instruction were (a) measurable benchmarks tied to grade-level content standards and (b) self-determination (the level of the student's self-direction within his or her own learning activities). Again these findings are consistent with those of Kampfer et al. (2001) and Karvonen et al. (2006) in emphasizing the importance of measurable targeted skills, integration of assessment and instruction, and opportunities for student self-direction.

Together, these seven studies suggest that teachers need considerably more training, as well as explicit examples and ongoing support, in making the connection between alternate assessment and daily instruction. If one of the key purposes of large-scale assessment within educational reform is to change how teachers teach, then clearly much work needs to be done if alternate assessments are to attain the promised impact on programs for students with significant cognitive disabilities.

Recommendation 5: Use Best Measurement Practice for Scoring and Reporting Alternate Assessments and Collecting and Reporting Data on Technical Quality. NCLB requires that alternate

assessments be technically valid instruments (NCLB, 2001; U.S. Department of Education, 2004). This includes being reliable as well as valid. Yet despite this requirement, scholars do not have a great deal of research to date that documents the technical adequacy of alternate assessments. Researchers are certainly beginning to document the consequences of AA-AAS from various stakeholder perspectives, but much work needs to be done in this area.

Consequential validity. A key component in addressing the validity of an assessment is consequential validity, or the consequences—both intended and unintended—of that assessment on student instruction and learning (Linn, Baker, & Dunbar, 1991; Marion & Pellegrino, 2006; Shepard, 1993). In essence, consequential validity answers the “So what” question for an assessment system. If the consequences of an assessment system are valid for the intended purposes and uses of that system, then the question (i.e., So what?) can be answered with confidence. As a measure of consequential validity, Kleinert, Kennedy, and Kearns (1999) completed a statewide teacher survey in Kentucky to determine their perceptions of the benefit of the AA-AAS on their students and schools. Overall, teachers noted the benefit of positive instructional changes as well as improved student outcomes. However, teachers noted their frustration with the amount of time necessary to complete the assessments and concerns with reliability of scoring.

Roach, Elliott, and Berndt (2007) also investigated variables that influenced teachers' perceptions of one state's AA-AAS. Overall, teachers reported ambivalence to the AA-AAS process, the usefulness of the assessment in monitoring students' learning, and the identification of what was most important to teach this population. Although teachers agreed that assessment items aligned well with the general academic standards, they were less positive in their overall perceptions of the alternate assessment for older, especially secondary level, students.

In a similar vein, Roach (2006) investigated the variables that influenced parents' perceptions of one state's AA-AAS. Results suggested that (a) parents believed the assessment to be useful; (b) that they found utility in their students learning academic content in reading, writing, and math;

and (c) that they generally agreed that all students should have the opportunity to participate in statewide assessments. Parents expressed their strongest agreement in the confidence in the results of their own child's performance (a mean of 4.25 on a 5-point Likert scale). However, parent perception was again less positive as a function of student age (as students got older) and as the ratio of functional to academic objectives on their child's IEP increased. From a parent perspective, the alternate assessment for that state would thus appear to be a less valid assessment for students whose school program was more focused on functional and independent living goals.

Towles-Reeves, Kampfer-Bohach, Garrett, Kearns, and Grisham-Brown (2006) noted that statewide coordinators for the deaf-blind have had little opportunity to be involved in the development of large-scale assessments within their respective states, as well as uncertainty among state coordinators for the deaf-blind regarding how these students were even performing in large-scale assessment systems. These findings suggest a need for collaboration among schools and coordinators for the deaf-blind at a deeper level in order to measure progress and improve outcomes for students with deaf-blindness participating in alternate assessments. Clearly, if key stakeholders are somehow absent from the entire process, it will be difficult for an assessment to produce its intended consequences.

Although many studies cite the negative consequences of assessment systems (Cizek, 2001), Ysseldyke, Dennison, and Nelson (2004) investigated the positive consequences of AA-AAS for students with significant cognitive disabilities. Four positive consequences were found: "increased participation of students in testing programs, higher expectations and standards, improved instruction, and improved performance" (p. 1). Overall, the findings suggest that both intended and unintended positive consequences of AA-AAS are occurring. However, research is necessary to understand how to develop assessments that consider avoidance of unintended negative consequences and promote the intended positive consequences.

Additional elements of technical adequacy. Consequential validity is but one element of the technical quality of an alternate assessment; cer-

tainly other elements of validity (e.g., content, construct, predictive); the overall reliability of the assessment; interrater scoring agreement; and the extent to which the assessment can be scaled up or related to the general assessment are all important aspects of technical adequacy.

Kleinert, Kearns, and Kennedy (1997) outlined the development of the Kentucky alternate assessment and reported data on initial implementation that reflected reliability, validity, and instructional impact results. At the time of this study in 1997, initial results were encouraging, though the authors did note considerable difficulty with interrater scoring reliability. In these early years, little was known about the technical aspects of AA-AAS systems.

[R]esearch is necessary to understand how to develop assessments that consider avoidance of unintended negative consequences and promote the intended positive consequences.

Tindal et al. (2003) examined the technical adequacy of one state's performance event alternate assessment in reading and math. The assessment was based on curriculum-based measures in these subject areas extended in their application to students with moderate and severe intellectual disabilities. Tindal et al. found general support for the technical adequacy of that state's assessment. However, it is important to note that 75 of the 437 participating students (17.6%) in the Tindal et al. study were not able to be assessed due to "extremely low academic skills" (p. 485) and though the alternate assessment was based on state standards for reading and math, these standards were taken from the kindergarten through second grades, even though participating students ranged from kindergarten to postsecondary-age students. The authors concluded that students with significant disabilities need to be assessed so that performance and progress can be documented on tasks reflecting construct validity while at the same time linked to the standards for all students.

In a subsequent study, Yovanoff and Tindal (2007) investigated "hypotheses pertaining to the technical adequacy and vertical scale alignment of Oregon's early reading alternate assessment with

the general education grade-level assessments" (p. 3). These authors argued that carefully selected, scaled performance tasks would provide an appropriate assessment option for students with significant cognitive disabilities. Specifically, the performance tasks were found to be technically appropriate for use in the alternate assessment for third-grade reading and to accurately assess these students' abilities; further, these items could be scaled in terms of a statewide testing metric aligned to the grade-level academic standards for all students. This approach would enable educators and policy makers to better understand the relationship of alternate assessment scores to general assessment scores.

In contrast, Johnson and Arnold (2004) found deficiencies in the "evidence for content, response process, and structural validity" (p. 266) of Washington State's alternate assessment portfolio system (WAAS). Consistent with other scholars (Browder, Ahlgrim-Delzell et al., 2005; Tindal et al., 2003), the authors suggested that much work is needed to establish an alternate assessment that reflects both adequate psychometric properties and instructional relevance for students with severe disabilities. Johnson and Arnold also noted that teachers' ability to correctly assemble student portfolios significantly impacted student scores; this finding was echoed in a subsequent study (Johnson & Arnold, 2007) in which these researchers found that a "large percentage of portfolios (78%–90%) may have received low scores because not all of the critical elements were included" (p. 26). This calls into question whether the alternate assessment was really a measure of teacher, rather than student, performance. Johnson and Arnold (2007) further noted that the IEP skills that teachers identified as exemplifying state standards for the assessment portfolios were so diverse that reasonable comparisons of exactly what scores meant were not possible.

Finally, Kohl et al. (2006) noted that 9 of 16 states in their study indicated that they had conducted validity and alignment studies of their respective alternate assessments. However, these authors could find little specific information about the majority of these studies, with most appearing to have been conducted by stakeholder groups on the extent to which the alternate assessment represented general state standards. Truly

rigorous studies of technical adequacy appeared to be missing.

Taken together, these studies suggest that stakeholders have a great deal to learn about the technical qualities of current state alternate assessments, and what they do know suggests insufficient technical adequacy or lack of technical documentation in most cases. Even more challenging for practitioners and policy makers will be the establishment of technical adequacy of alternate assessments for students with the most significant cognitive disabilities that are able to appropriately assess this heterogeneous group of students and that are linked to grade-level content standards. Although the work of Yovanoff and Tindall (2007) is promising for early-grade reading tasks, for students at upper elementary, middle, and high school ages, establishing the technical adequacy of this linkage may prove progressively difficult. Future work in developing technically sound assessments is clearly vital to meet the requirements of NCLB.

Recommendation 6: Use Alternate Assessment Outcomes for Program Evaluation and Ongoing Quality Enhancement. Three studies have investigated the relationship or use of alternate assessment outcomes for program evaluation. To investigate life outcomes for students in the AA-AAS, Kleinert et al. (2002) interviewed students who had exited school 1 year earlier and compared their postschool outcomes with their alternate assessment scores. Results revealed no connection between these students' alternate assessment scores and their postschool outcomes. An important finding, though, was that students' lack of verbal communication skills was invariably indicative of poor postschool outcomes. This finding is especially critical in that students with limited or nonexistent verbal skills make up a significant portion of students in state AA-AAS (Towles-Reeves, Kearns, Kleinert, & Kleinert, 2008).

Karvonen et al. (2006) found that resources, teacher characteristics, and instructional effectiveness are important factors that contribute to alternate assessment outcomes. This year-long, qualitative study examined the factors that contributed to alternate assessment outcomes among seven teachers and their students with significant cognitive disabilities who participated in their

state's alternate assessment. Consistency across cases revealed the strength or quality of instructional programs and the effort of teachers in developing high-quality portfolios as important predictors of student scores.

Browder, Spooner, Algozzine, et al. (2003) noted that "one of the hoped-for impacts of alternate assessment is that it will improve educational outcomes for students with significant cognitive disabilities" (p. 57). Roach and Elliott (2006) investigated this proposition by studying the influence of access to the general curriculum on students' performance on the Wisconsin Alternate Assessment. Results indicated that students who had access to the general curriculum also performed better on the alternate assessment in the domains of reading, language arts, and mathematics.

For alternate assessments to have utility in improving overall program quality, and most importantly outcomes for students, several elements would need to be embedded into the alternate assessment process. First, there is a need for the provision of ongoing targeted professional development to help teachers conduct alternate assessments that are themselves of high quality. In particular, states will need to make efforts to equip teachers with new resources to help them be effective in meeting the demands of NCLB and to help students with the most significant cognitive disabilities meet state expectations. As noted by Karvonen et al. (2006) and Roach and Elliott (2006), this is not simply a matter of professional development and resources directed to assessment but rather training and resources to improve ongoing instruction (e.g., direct instruction, data-based decision-making, IEP objectives with clear links to state standards, access to the general curriculum, and student-directed learning).

Finally, if alternate assessments are to be viewed as useful tools for program evaluation, then there must be evidence that how students do on alternate assessments relates to other measures of student outcomes. Researchers need more than just teacher reports that alternate assessments result in improved student outcomes (Johnson & Arnold, 2007); they need data-based studies that reflect clear gains in both academics and other critical life outcomes.

ALTERNATE ASSESSMENTS BASED ON ALTERNATE ACHIEVEMENT STANDARDS— AN EVOLVING PICTURE

Thus far we have presented a review of alternate assessment literature in terms of the six specific recommendations of Browder, Spooner, Algozzine, et al. (2003). However, it is also important to note the relative infancy of this field (alternate assessments were first mandated under IDEA for all states starting July 2000) and how even in this brief period the speed with which the field has evolved. The National Center on Educational Outcomes (NCEO) has summarized results of surveys of State Directors of Special Education every 2 years from 1999 through 2005 with an online survey created in 2000 to record the status of states as the AA-AAS requirements were put in place and another online survey regarding states' development of AA-AAS (Thompson, Erickson, Thurlow, Ysseldyke, & Callender, 1999; Thompson & Thurlow, 1999, 2000, 2001, 2003; Thompson et al., 2005). The purpose of these reports was to make public the trends and issues facing states, as well as the innovations that states are using to meet the demands of federal legislation. These reports have illuminated the changes that have occurred across states in the development of their AA-AAS, given very significant changes in federal policy and research related to improving education and assessment for students with the most significant cognitive disabilities. Thompson and colleagues have noted that, although there has been progress of students with disabilities toward proficiency in this era of standards-based accountability, states continue to face many challenges in their efforts to increase student achievement and administer assessments that provide valid documentation of this achievement.

For example, Thompson and Thurlow (2003) found in their survey of 61 state directors in special education (including unique states) "more students with disabilities are accessing state/district academic content standards with increased academic expectations, and increased participation in accountability systems" (p. 4). Similarly, Thompson et al. (2005) noted that states reported an increase in the total number of students with disabilities achieving proficiency on state accountability tests (i.e., all students with disabilities, not

only students taking alternate assessments). States also reported continued growth in understanding the approach, content, standard setting, and scoring criteria of alternate assessments (Thompson et al., 2005). The authors suggested that although there has been marked progress, some difficult issues still remain, including (a) the need for re-designing current state alternate assessments to reflect grade-level academic content, (b) standardization and improved technical quality of alternate assessments, (c) standard setting, and (d) administration issues such as time and cost.

NCEO has also conducted an analysis of state accommodations policies and guidelines every 2 years since 2001 (Clapper, Morse, Lazarus, Thompson, & Thurlow, 2005; Lazarus, Thurlow, Kail, Eisenbraun, & Kato, 2006; Thurlow, Lazarus, Thompson, & Robey, 2002). Results across these reports suggested that states were continuing "to adjust their policies to ensure that students with disabilities have opportunities to participate in statewide assessments, and at the same time to understand the meaning of the scores from their assessments" (Thurlow et al., p. 2). However, the intent of these reports was only to provide a descriptive analysis (not quality analysis) of the policies put forth by states related to accommodations and assessment.

QUALITY AND LIMITATIONS OF RESEARCH ON ALTERNATE ASSESSMENTS

Although we did not attempt to evaluate the extent to which the studies reviewed in this article embody the Council for Exceptional Children criteria for evidence-based practices (Graham, 2005), there are some observations that can be made about research on AA-AAS as a whole. First, much of the research conducted thus far has fallen into two major lines of inquiry:

1. Teacher perceptions of instructional impact (Flowers et al., 2005; Kampfer et al., 2001; Kleinert et al., 1999; Roach et al., 2007; Towles-Reeves et al., 2006; Towles-Reeves & Kleinert, 2006); parent perceptions (Roach, 2006); or severe disability expert perceptions (Kleinert & Kearns, 1999) of the AA-AAS process.
2. The extent to which the performance indicators for students in the AA-AAS are reflective of the content standards established for all students (Browder, Ahlgrim-Delzell, et al., 2005; Browder et al., 2004; Browder, Spooner, Ahlgrim-Delzell, et al., 2003; Flowers, Browder, & Ahlgrim-Delzell, 2006; Roach et al., 2005).

Certainly, these types of studies are important for understanding the extent to which alternate assessments do reflect academic content standards and the perceptions of the impact of alternate assessments by key stakeholders in the process. Yet, there is considerably less research that has examined the extent to which actual student scores were associated with empirically verified instructional or other outcome variables (Karvonen et al., 2006; Kleinert et al., 2002; Turner et al., 2000; White et al., 2003), and only one study thus far that employed an experimental and a control group to systematically examine the impact of a teacher intervention on alternate assessment scores and the achievement of IEP objectives (Browder, Karvonen, et al., 2005). Moreover, though approaches to documenting technical adequacy of student AA-AAS measures are promising (Tindal et al., 2003; Yovanoff & Tindal, 2007), this work has yet to be extended to upper grade students or to students with severe and multiple disabilities.

In short, there is a need not only for research along the critical lines that we outline below, but for a second generation of AA-AAS research that more rigorously examines, including experimentally controlled studies whenever possible, the questions we posit. Moreover, this second generation of research must include, in its sampling of students, the full range of students who participate in AA-AAS. It is only in pursuing such a second generation of research that we can clearly identify and describe evidence-based practices for AA-AAS.

FUTURE DIRECTIONS AND POLICY DECISIONS REGARDING ALTERNATE ASSESSMENT

Clearly, the six categories for research suggested by Browder, Spooner, Algozzine, et al. (2003)

remain relevant today. As a field, it would appear that researchers have made at least limited gains in five of the six areas of inquiry for alternate assessments suggested by Browder, Spooner, Algozine, et al. At the same time, especially given the impact of NCLB on school, district, and state accountability, there are additional areas of research that must be considered to further our understanding of the impact of AA-AAS.

CONSIDERATIONS FOR ADDITIONAL RESEARCH

First, future research should include a systematic investigation and description of the population of learners who participate in AA-AAS, as well as the explication of a theory of learning for this population. These students are an extremely heterogeneous population that has unique and specialized support needs to learn grade-level academic content. Yet, in the field researchers have not investigated (a) if there are specific characteristics about these students that are common across states and assessments (e.g., receptive-expressive language characteristics, vision-hearing characteristics, motor characteristics, etc.), or (b) if there are typical learning patterns or "cognitive maps" that may emerge from this group.

Typically learning students acquire and store knowledge related to reading and mathematics in certain ways (see Pellegrino, Chudowsky, & Glaser, 2001). Therefore, teachers target these learning styles in the classroom. For students with the most significant cognitive disabilities, researchers have yet to consider exactly how these students acquire, retain, and generalize academic knowledge. For example, students with significant cognitive disabilities may develop and build on knowledge schema related to academic content areas (e.g., reading, mathematics) in different ways than other students. At this point, researchers simply cannot answer the question of how students with significant cognitive disabilities represent knowledge and acquire domain competence across traditional academic subjects.

This may especially be the case for students with severe and profound intellectual disabilities. For example, Towles-Reeves et al., (2008) found, across three geographically and demographically distinct states, that a significant subset of this pop-

ulation has not yet acquired the use of fully symbolic modes of communication (i.e., the use of words, signs, or other means of formal language) but may still be at emerging symbolic (e.g., communication supported by picture and/or gestural prompts) or presymbolic (e.g., communication primarily through facial expressions, changes in muscle tone) communication levels. These authors found that approximately 8% to 11% of students in these three states' AA-AAS had language skills that could best be described as presymbolic, with even larger percentages of students in each of the three states having no observable awareness of print or Braille (15%, 25%, and 13%, respectively) and no observable awareness or use of numbers (13%, 22%, and 11%, respectively). How to design alternate assessments systems, especially alternate assessments linked to grade-level content standards, for students at emerging and presymbolic levels in a way that allows these students to demonstrate what they do know is an immense challenge for the field.

Our second recommendation for future research is to investigate the relationship between AA-AAS (regardless of approach: portfolios, performance assessments, and checklists) and another accepted measure of student learning. Most tests used in the educational setting such as intelligence, achievement, or adaptive measures support their validity by evidencing correlations with other measures that purport to measure the same construct. The unique nature of AA-AAS makes this difficult. AA-AAS are based on individual states' academic content standards making comparisons inappropriate with other educational tests (e.g., intelligence, adaptive measures) that are not based on these standards. Although Browder, Karvonen, et al. (2005) did find a correlation between improved alternate assessment scores and achievement of student IEP objectives, there is no evidence to support the correlation of alternate assessments with other accepted measures of student learning (e.g., IEP reviews). Finally, for students taking the alternate assessment in their final high school years, one would expect at least some correlation with postschool outcomes for all students in the alternate assessment. The field will have an increased capability to do this type of research, as states put into place the student postschool outcomes data sets required by IDEA 2004.

Although there is evidence that increasing numbers of students taking alternate assessments are achieving proficiency (Thompson et al., 2005), does this trend toward proficiency represent enhanced student learning or simply the fact that teachers have become more experienced with the requirements of the alternate assessment? In this sense and others, the technical adequacies of alternate assessments are still in question.

Finally, a commonly accepted, conceptual framework for evaluating the validity of AA-AAS may be of value to both policy makers and practitioners. Marion and Pellegrino (2006) have proposed such a framework. Their model places consequential validity as the central validity benchmark for this type of assessment while also addressing other essential elements of technical quality related to AA-AAS.

CONSIDERATIONS DUE TO CHANGES IN FEDERAL POLICY

An additional area of research that must be considered is how well alternate assessments are meeting new federal legislation that requires AA-AAS to be linked to grade-level academic content standards (U.S. Department of Education, 2004, 2005a). In the past, U.S. Office of Special Education Program guidance (Heumann & Warlick, 2000) clarified that alternate assessments had to be based on the same academic standards as the general assessment, but this was a broadly worded policy guidance statement that did not describe specific grade-level expectations. Although alternate achievement standards may vary from grade-level achievement standards in depth, breadth, and complexity (U.S. Department of Education, 2004, 2005a), students with significant cognitive disabilities must receive instruction linked to grade-level academic content standards if they are to be assessed on achievement standards linked to that grade-level content. Because this requirement is so recent, practitioners and policy makers are unsure of its impact on student outcomes, instructional practices, or the number of students with significant cognitive disabilities who will score proficient on their respective state assessments as a result of this requirement.

At an even more basic level, the field needs to consider what constitutes "sufficient linkage" to

grade-level content standards for students with significant cognitive disabilities. Flowers et al. (2006) found that for three states judged to have exemplary alternate assessments, those states' alternate assessments sampled a restricted range of content standards with a reduced depth or complexity of knowledge. Clearly, in its guidance documents, the U.S. Department of Education (2005a) allows for this reduced depth and breadth in alternate assessments on alternate achievement standards, but at what point does this reduced scope and complexity result in the measurement of skills that have little semblance to the content standards to which these skills are supposedly linked?

A second federal policy change that will directly impact the course of future research and practice is the introduction of AA-GLAS and AA-MAS. Although a discussion of these two types of alternate assessment are beyond the scope of this review, the same questions that we have identified previously will have to be addressed for AA-GLAS and AA-MAS (e.g., defining the assessment population, a framework of learning for these students, technical adequacy of the instruments, what constitutes adequate alignment of content for grade-level or modified achievement standards). The reader is referred to the following documents for further information about AA-GLAS: U.S. Department of Education's (2005b) *Tool Kit on Teaching and Assessing Students With Disabilities* as well as NCEO Synthesis Report 59 (Wiener, 2006). For further information about AA-MAS, readers are referred to the Non-Regulatory Guidance on AA-MAS (U.S. Department of Education, 2007).

IMPLICATIONS FOR RESEARCHERS AND POLICY MAKERS

Although the studies cited previously have informed the field of alternate assessment, there is still not enough data-based evidence to say with confidence that alternate assessment has achieved the promises articulated by Browder, Spooner, Algozzine, et al. (2003): (a) greater consideration of students with disabilities in school and state policy decisions, (b) overall increased expectations for individuals with disabilities, (c) increased access to

the general curriculum and state academic content standards for these students, and (d) improvement of instructional decision making at both the teacher and classroom level. We suggest the following research agenda as critical to the field in the years ahead in documenting those promises:

1. Studies of technical adequacy of alternate assessments will become increasingly important in an era of high-stakes accountability for both schools and students. The limited research available on the technical adequacy of alternate assessments has yet to consider the requirement that AA-AAS must now be linked to grade-level content standards. Measurement, core academic content, and severe disability experts have recently begun to work together to develop systematic approaches or frameworks for technical adequacy for alternate assessments (see Marion & Pellegrino, 2006) and to pilot these frameworks within a representative sample of states (National Alternate Assessment Center, 2007; see <http://www.naacpartners.org/>). However, this work is still in its infancy but should be considered when conducting research related to the technical adequacy of AA-AAS.
2. We need to clearly document the extent to which alternate assessments are improving access to the general curriculum, even for students with the most significant cognitive disabilities. Teachers need research-based examples of how they can provide such access while still providing embedded functional skill instruction on those more basic skills that many students with significant disabilities require for enhanced independence and community participation.
3. It is not enough for states to establish trend lines illustrating that increasing numbers of students are achieving proficiency on their respective state alternate assessments. Rather, researchers must show how improved alternate assessment scores correlate with other valued measures of student learning and outcomes (e.g., curriculum-based assessments and postschool outcomes). The one study in the literature that we do have on the relationship of alternate assessment scores to postschool outcomes (see Kleinert et al., 2002)

did not find a relationship between alternate assessment scores and postschool student outcomes. Clearly, without such evidence of enhanced student outcomes, it is hard to justify the increased teacher workload of alternate assessments.

IMPLICATIONS FOR PRACTITIONERS

There are also significant issues for practitioners in the implementation of AA-AAS. In this final section, we consider not only the practical issues that teachers and administrators face in implementing these assessments but also implications for parents and students as well.

Linkage to grade-level content standards for students with significant cognitive disabilities will present considerable challenges for practitioners. As we have previously noted, we must determine what constitutes sufficient linkage, and in many cases, this decision may fall to teachers and IEP teams. For example, what does linkage look like for students who do not yet have symbolic modes of communication? Although states have the obligation to place these alternate assessments into practice now (and teachers have the obligation to implement these assessments with their students with significant cognitive disabilities), researchers and policy makers, working in close collaboration with practitioners, have the obligation to develop guidelines that result in both defensible linkages to grade-level content standards and instruction in meaningful skills for students with severe disabilities. Part of this work includes a clear description of who the learners are and in what ways these students' methods of constructing academic knowledge are similar or dissimilar from typical learners.

Most teachers believe that students with the most significant cognitive disabilities deserve to be part of school and district accountability systems (see Flowers et al., 2005; Kleinert et al., 1999) and researchers have not really documented the broader administrative or policy consequences to these students from that involvement. For example, how are alternate assessments scores used, if at all, by districts and schools in formulating district and school improvement plans? Do the results of alternate assessments really make a difference in the classroom and school-level prac-

tices? Are school and district administrators, as well as general educators, more supportive of the inclusion of students with significant cognitive disabilities in grade-level curricular activities?

Finally, there are two sets of voices conspicuously absent in the research on alternate assessment conducted thus far—those of parents and of the students themselves. We could find only one study to date that systematically sampled parental perceptions of the impact of alternate assessments on the education of their children (Roach, 2006), and no studies that asked the students who participate in alternate assessments if these assessments have helped them to learn, have resulted in increased control or decision-making over their own learning, or have enabled them to understand what they have accomplished in school. Clearly, stakeholders need to understand the perspectives of families and students if they are to fully understand the consequences of alternate assessments.

CONCLUSION

This review has integrated the data-based studies on AA-AAS conducted since the conception of AA-AAS. Although we have found limited progress in several elements of the research framework proposed by Browder et al. (2003; e.g., validating performance indicators with content area experts and stakeholders, training teachers in incorporating alternate assessments into daily practice), very significant gaps remain in scholars' ability to provide practitioners and policy makers with research-based strategies that will enable alternate assessment to truly achieve its promises to students, teachers, and parents.

REFERENCES

Browder, D., Ahlgrim-Delzell, L., Flowers, C., Karvonen, M., Spooner, F., & Algozzine, R. (2005). How states implement alternate assessments for students with disabilities. *Journal of Disability Policy Studies, 15*(4), 209–220.

Browder, D., Flowers, C., Ahlgrim-Delzell, L., Karvonen, M., Spooner, F., & Algozzine, R. (2004). The alignment of alternate assessment content with academic and functional curricula. *The Journal of Special Education, 37*(4), 211–223.

Browder, D., Karvonen, M., Davis, S., Fallin, K., & Courtade-Little, G. (2005). The impact of teacher training on state alternate assessment scores. *Exceptional Children, 71*, 267–282.

Browder, D., Spooner, F., Ahlgrim-Delzell, L., Flowers, C., Algozzine, R., & Karvonen, M. (2003). A content analysis of the curricular philosophies reflected in states' alternate assessment performance indicators. *Research and Practice for Persons With Severe Disabilities, 28*(4), 165–181.

Browder, D., Spooner, F., Ahlgrim-Delzell, L., Harris, A., & Wakeman, S. (2008). A meta-analysis on teaching mathematics to students with significant cognitive disabilities. *Exceptional Children, 74*, 407–432.

Browder, D., Wakeman, S., Flowers, C., Rickelman, R., Pugalee, D., & Karvonen, M. (2007). Creating access to the general curriculum with links to grade-level content for students with significant cognitive disabilities: An explication of the concept. *The Journal of Special Education, 41*(1), 2–16.

Browder, D., Wakeman, S., Spooner, F., Ahlgrim-Delzell, L., & Algozzine, B. (2006). Research on reading instruction for individuals with significant cognitive disabilities. *Exceptional Children, 72*, 392–408.

Browder, D. M., Spooner, R., Algozzine, R., Ahlgrim-Delzell, L., Flowers, C., & Karvonen, M. (2003). What we know and need to know about alternate assessment. *Exceptional Children, 70*, 45–61.

Cizek, G. (2001). Conjectures on the rise and fall of standard setting: An introduction to context and practice. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 3–17). Mahwah, NJ: Erlbaum.

Clapper, A. T., Morse, A. B., Lazarus, S. S., Thompson, S. J., & Thurlow, M. L. (2005). *2003 state policies on assessment participation and accommodations for students with disabilities* (Synthesis Report 56). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved August 9, 2007, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis56.html>

Courtade, G., Spooner, F., & Browder, D. (2007). A review of studies with students with significant cognitive disabilities that link to science standards. *Research and Practice in Severe Disabilities, 32*, 43–49.

Destefano, L., Shriner, J., & Lloyd, C. (2001). Teacher decision making in participation of students with disabilities in large-scale assessment. *Exceptional Children, 68*, 7–22.

- Farlow, L. J., & Snell, M. E. (1994). *Making the most of student performance data. (Innovations)*. Washington, DC: American Association on Mental Retardation.
- Flowers, C., Ahlgrim-Delzell, L., Browder, D., & Spooner, F. (2005). Teachers' perceptions of alternate assessments. *Research and Practice for Persons With Severe Disabilities, 30*(2), 81–92.
- Flowers, C., Browder, D. M., & Ahlgrim-Delzell, L. (2006). An analysis of three states' alignment between language arts and mathematics standards and alternate assessments. *Exceptional Children, 72*, 201–215.
- Flowers, C., Wakeman, S., Browder, D., & Karvonen, M. (2007). *Links for academic learning: An alignment protocol for alternate assessments based on alternate achievement standards*. Charlotte, NC: National Alternate Assessment Center, University of North Carolina at Charlotte.
- Graham, S. (Ed.). (2005). Criteria for evidenced-based practice in special education [Special issue]. *Exceptional Children, 71*, 137–207.
- Heumann, J., & Warlick, K. (2000, August 24). *OSEP memorandum to state directors of special education (OSEP 00-24)*. Washington, DC: U.S. Department of Education.
- Horvath, L., Kampfer-Bohach, S., & Kearns, J. (2005). The use of accommodations among students with deaf-blindness in large-scale assessment systems. *Journal of Disability Policy Studies, 16*(3), 177–187.
- Individuals With Disabilities Education Act Amendments of 1997 (IDEA), Pub. L. No. 105-17, 20 U.S.C. §§ 1400 *et seq.*
- Individuals With Disabilities Education Improvement Act of 2004 (IDEA), Pub. L. No. 108-446, 20 U.S.C. §§ 1400 *et seq.*
- Johnson, E., & Arnold, N. (2004). Validating an alternate assessment. *Remedial and Special Education, 25*(5), 266–275.
- Johnson, E., & Arnold, N. (2007). Examining an alternate assessment: What are we testing? *Journal of Disabilities Studies, 18*(1), 23–31.
- Kampfer, S., Horvath, L., Kleinert, H., & Kearns, J. (2001). Teachers' perceptions of one state's alternate assessment portfolio program: Implications for practice and preparation. *Exceptional Children, 67*, 361–374.
- Karvonen, M., Flowers, C., Browder, D., Wakeman, S., & Algozzine, B. (2006). Case study of the influence on alternate assessment outcomes for students with disabilities. *Education and Training in Developmental Disabilities, 41*(2), 95–110.
- Karvonen, M., Wakeman, S., Flowers, C., & Browder, D. (2007). Measuring the enacted curriculum for students with significant cognitive disabilities. *Assessment for Effective Intervention, 33*, 29–38.
- Kleinert, H., Garrett, B., Towles, E., Garrett, M., Nowak-Drabik, K., Waddell, C., et al. (2002). Alternate assessment scores and life outcomes for students with significant disabilities: Are they related? *Assessment for Effective Intervention, 28*, 19–30.
- Kleinert, H., & Kearns, J. (2004). Alternate assessments. In F. Orellove, D. Sobsey, & R. Silberman (Eds.), *Educating children with multiple disabilities: A collaborative approach* (4th ed., pp. 115–149). Baltimore: Paul Brookes.
- Kleinert, H., Kearns, J., & Kennedy, S. (1997). Accountability for all students: Kentucky's Alternate Portfolio system for students with moderate and severe cognitive disabilities. *Journal of the Association for Persons With Severe Handicaps, 22*(2), 88–101.
- Kleinert, H., & Kearns, J. F. (1999). A validation study of the performance indicators and learner outcomes of Kentucky's alternate assessment for students with significant disabilities. *Journal of the Association for Persons With Severe Handicaps, 24*, 100–110.
- Kleinert, H., Kennedy, S., & Kearns, J. (1999). Impact of alternate assessments: A statewide teacher survey. *Journal of Special Education, 33*(2), 93–102.
- Kohl, F., McLaughlin, M., & Nagle, K. (2006). Alternate achievement standards and assessments: A descriptive investigation of 16 states. *Exceptional Children, 73*, 107–123.
- Lazarus, S., Thurlow, M., Kail, K., Eisenbraun, K., & Kato, K. (2006). *2005 state policies on assessment participation and accommodations for students with disabilities (Synthesis Report 64)*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved August 9, 2007, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis64/>
- Linn, R., Baker, E., & Dunbar, S. (1991). Complex, performance based assessment: Expectations and validation criteria. *Educational Researcher, 20*, 15–21.
- Marion, S., & Pellegrino, J. (2006). A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practices, 25*(4), 47–57.
- National Alternate Assessment Center. (2007). *Technical documentation workbooks I and II*. Lexington: University of Kentucky, National Alternate Assessment Center. Retrieved August 1, 2007, from <http://www.naacpartners.org/products.aspx>
- No Child Left Behind Act of 2001, Pub. L. 107-110, 115 Stat. 1425, 20 U.S.C. §§ 6301 *et seq.*

- Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: Committee on the Foundations of Assessment, National Academy Press.
- Perner, D. (2007). No child left behind: Issues of assessing children with the most significant cognitive disabilities. *Education and Training in Developmental Disabilities, 42*, 243–251.
- Quenemoen, R., Rigney, S., & Thurlow, M. (2002). *Use of alternate assessment results in reporting and accountability systems: Conditions for use based on research and practice* (Synthesis Report 43). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved April 16, 2006, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis43.html>
- Roach, A. (2006). Influences on parent perceptions of an alternate assessment for students with severe cognitive disabilities. *Research & Practice for Persons With Severe Disabilities, 31*, 267–274.
- Roach, A., & Elliott, S. (2006). The influences of access to general education curriculum on alternate assessment performance of students with significant cognitive disabilities. *Educational Evaluation and Policy Analysis, 28*(2), 181–194.
- Roach, A., Elliott, S., & Berndt, S. (2007). Teacher perceptions and the consequential validity of an alternate assessment for students with significant disabilities. *Journal of Disability Policy Studies, 18*, 168–175.
- Roach, A., Elliott, S., & Webb, N. (2005). Alignment of alternate assessment with state academic standards: Evidence for the content validity of the Wisconsin alternate assessment. *Journal of Special Education, 38*(4), 218–231.
- Roeber, E. (2002). *Setting standards on alternate assessments* (Synthesis Report 42). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved November 28, 2005, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis42.html>
- Shepard, L. A. (1993). Evaluating test validity. In L. Darling-Hammond (Ed.), *Review of research in education* (Vol. 19, pp. 405–450). Washington, DC: American Educational Research Association.
- Thompson, S., Erickson, R., Thurlow, M., Ysseldyke, J., & Callender, S. (1999). *Status of the states in the development of alternate assessments* (Synthesis Report 31). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved November 25, 2005, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis31.html>
- Thompson, S., Johnstone, C., Thurlow, M., & Altman, J. (2005). *2005 State special education outcomes: Steps forward in a decade of change*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved December 12, 2005, from <http://www.education.umn.edu/nceo/OnlinePubs/2005StateReport.htm>
- Thompson, S., & Thurlow, M. (1999). *1999 State special education outcomes: A report on state activities at the end of the century*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved November 28, 2005, from <http://education.umn.edu/NCEO/OnlinePubs/99StateReport.htm>
- Thompson, S., & Thurlow, M. (2000). *State alternate assessments: Status as IDEA alternate assessment requirements take effect* (Synthesis Report 35). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved November 28, 2005, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis35.html>
- Thompson, S., & Thurlow, M. (2001). *2001 State special education outcomes: A report on state activities at the beginning of a new decade*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved November 28, 2005, from <http://education.umn.edu/NCEO/OnlinePubs/2001StateReport.html>
- Thompson, S., & Thurlow, M. (2003). *2003 State special education outcomes: Marching on*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved November 28, 2005, from <http://education.umn.edu/NCEO/OnlinePubs/2003StateReport.htm>
- Thompson, S., Thurlow, M., Esler, A., & Whetstone, P. (2001). Addressing standards and assessments on the IEP. *Assessment for Effective Intervention, 26*(2), 77–84.
- Thurlow, M., Lazarus, S., Thompson, S., & Robey, J. (2002). *2001 state policies on assessment participation and accommodations* (Synthesis Report 46). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved December 12, 2005, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis46.html>
- Tindal, G., McDonald, M., Tedesco, M., Glasgow, A., Almond, P., Crawford, L., et al. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Exceptional Children, 69*, 481–494.
- Towles-Reeves, E., Garrett, B., Burdette, P., & Burdige, M. (2006). What are the consequences? Validation of large-scale alternate assessment systems and their influence on instruction. *Assessment for Effective Intervention, 31*(3), 45–57.

- Towles-Reeves, E., Kampfer-Bohach, S., Garrett, B., Kearns, J., & Grisham-Brown, J. (2006). Are we leaving our children behind? State deafblind coordinators' perceptions of large-scale assessment. *Journal of Disability Policy Studies, 17*(1), 40–47.
- Towles-Reeves, E., Kearns, J., Kleinert, H., & Kleinert, J. (2008, May 12). An analysis of the learning characteristics of students taking alternate assessments based on alternate achievement standards. *Journal of Special Education*. Retrieved June 2, 2008, from <http://sed.sagepub.com/cgi/rapidpdf/0022466907313451v1>.
- Towles-Reeves, E., & Kleinert, H. (2006). The impact of one state's alternate assessment upon instruction and IEP development. *Rural Special Education Quarterly, 25*(3), 31–39.
- Turner, M., Baldwin, L., Kleinert, H., & Kearns, J. (2000). An examination of the concurrent validity of Kentucky's alternate assessment system. *Journal of Special Education, 34*(2), 69–76.
- U.S. Department of Education. (2003). *Education week analysis of data from the Office of Special Education Programs, Data Analysis System*. Washington, DC: Author.
- U.S. Department of Education. (2004). *Standards and assessment peer review guidance*. Washington, DC: Office of Elementary and Secondary Education.
- U.S. Department of Education. (2005a). *Alternate achievement standards for students with the most significant cognitive disabilities: Non regulatory guidance*. Washington, DC: Office of Elementary and Secondary Education.
- U.S. Department of Education. (2005b). *The tool kit on teaching and assessing students with disabilities*. Washington, DC: Office of Special Education. Retrieved August 30, 2006, from <http://www.osepideasthatwork.org/toolkit/index.asp>
- U.S. Department of Education. (2007). *Modified academic achievement standards: Non-regulatory guidance*. Washington, DC: Author. Retrieved July 30, 2007, from <http://www.ed.gov/policy/speced/guid/modachieve-summary.html>
- Webb, N. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. (NISE Research Monograph No. 6). Madison: University of Wisconsin—Madison, National Institute for Science Education.
- Webb, N. (2002, April). *An analysis of the alignment between mathematics standards and assessments for three states*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- White, M., Garrett, B., Kearns, J., & Grisham-Brown, J. (2003). Instruction and assessment: How students with deaf-blindness fare in large-scale alternate assessments. *Research and Practice for Persons With Severe Disabilities, 28*(4), 205–213.
- Wiener, D. (2006). *Alternate assessments measured against grade-level achievement standards: The Massachusetts "Competency Portfolio"* (Synthesis Report 59). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved November 30, 2006, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis59.html>
- Yovanoff, P., & Tindal, G. (2007). Scaling early reading alternate assessments with statewide measures. *Exceptional Children, 73*, 202–223.
- Ysseldyke, J., Dennison, A., & Nelson, R. (2004). *Large-scale assessment and accountability systems: Positive consequences for students with disabilities* (Synthesis Report 51). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved September 27, 2005, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis51.html>

ABOUT THE AUTHORS

ELIZABETH TOWLES-REEVES (CEC KY Federation), Research Coordinator, National Alternate Assessment Center; **HAROLD KLEINERT** (CEC KY Federation), Executive Director, Interdisciplinary Human Development Institute; and **MONICAH MUHOMBA**, Student, Educational and Counseling Psychology, University of Kentucky, Lexington.

Address correspondence to Elizabeth Towles-Reeves, National Alternate Assessment Center, University of Kentucky, 1 Quality St., Suite 722, Lexington, KY 40507. Phone: 859-257-7672 ext. 80255. Fax: 859-323-1838 (e-mail: liztowles-reeves@uky.edu).

This manuscript was supported, in part, by the U.S. Department of Education Office of Special Education Programs (Grant No. H3244040001). However, the opinions expressed do not necessarily reflect the position or policy of the U.S. Office of Special Education Programs and no official endorsement should be inferred.

Manuscript received April 2006; manuscript accepted October 2007.

Copyright of Exceptional Children is the property of Council for Exceptional Children and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.