Abstract: Council for Exceptional Children, Division on Developmental Disabilities (DDD), Position Paper approved by the DDD Board on January 31, 2007

Neither school reform nor the focus on developing accountability systems that measure and evaluate the success of schooling are new or innovative educational practices. Even before the No Child Left Behind (NCLB) Act of 2001, many states had changed or were changing their accountability and assessment systems. Similarly, individual school districts or local education authorities, often in the name of school reform, were using their own standards or benchmarks to assess progress for school improvement purposes and for fulfilling long-term district goals. However, with the NCLB mandate, state departments of education have had to specify what all students must achieve through state standards; and for students with the most significant cognitive disabilities, through state departments alternative achievement standards (U.S. State Department of Education, 2003). Many of the content standards, developed to date, continue to be expanded upon and/or refined. Likewise, states have developed or refined statewide assessments based on their own standards to show and monitor student and school achievement.

While the Division on Developmental Disabilities (DDD) supports accountability systems to evaluate school and student achievement and progress, as well as the use of statewide assessments based on achievement and content standards, we are concerned with how states determine accountability, and how they develop and implement assessments that affect students with disabilities. In particular, this DDD position paper will address a number of issues related to statewide assessments used to evaluate students with the most significant cognitive disabilities. The main issues that concern DDD are as follows: the variation of alternate assessments from state to state; the need to prove that the assessment instruments and procedures used are technically sound and appropriately identify proficient (and above) or non-proficient performance; and, the lack of awareness of or attention to the need for pre-service and in-service teacher training in assessing students with alternate assessments and using the results of the alternate assessments to guide ensuing instruction. These and other issues presented in this paper lead to the evaluation of how states are determined accountable for the achievement of students with the most significant cognitive disabilities. Additionally, many of the issues raised relate to all students who may not be able to achieve grade-level standards because of their special instructional needs.

Since the Individuals with Disabilities Education Act (IDEA) Amendments of 1997, states have been required to include students with disabilities in state and system-wide assessments. This includes students with significant cognitive disabilities. In addition, NCLB has identified the specific provisions for the alternate assessments required for students with the most significant disabilities (Flowers, Ahlgrim-Delzell, Browder, & Spooner, 2005). Alternate assessment has been defined by the U.S. Department of Education (2003) as:

Correspondence concerning this article should be addressed to Darlene E. Perner, Dept. of Exceptionality Programs, Bloomsburg University of Pennsylvania, 400 East Second Street, Bloomsburg, PA 17815.
An alternate assessment is an assessment designed for the small number of students with disabilities who are unable to participate in the regular State assessment, even with appropriate accommodations. An alternate assessment may include materials collected under several circumstances, including (1) teacher observation of the student, (2) samples of student work produced during regular classroom instruction that demonstrate mastery of specific instructional strategies in place of performance on a computer-scored multiple-choice test covering the same content and skills, or (3) standardized performance tasks produced in an “on demand” setting, such as completion of an assigned task on test day. To serve the purposes of assessment under title I, an alternate assessment must be aligned with the State’s content standards, must yield results separately in both reading/language arts and mathematics, and must be designed and implemented in a manner that supports use of the results as an indicator of AYP. (p. 68699)

For many states and local education authorities (LEAs), system-wide assessments based on content that reflect standards have become a way of school life. Although the requirements of NCLB have been outlined and are being implemented by state Departments of Education and LEAs, the U.S. Department of Education has allowed for flexibility in the design and implementation of these requirements. As a result, states and LEAs have been challenged with the responsibility of trying to meet all the requirements of NCLB and the many issues and decisions related to accountability, as well as student achievement and system-wide assessments. On January 14, 2004, the U.S. Department of Education identified 40 issues that states and LEAs have taken responsibility for in their systems design and implementation to meet average yearly progress (AYP) and the required 2014 timeline for all students to meet academic proficiency including students with the most significant cognitive disabilities. As a result, the design of accountability systems (including assessments) and the implementation of various provisions of NCLB have led to great variation from one state to another (see Table 1). In some states, this variation can also be from one district to another (e.g., Wyoming). This has made it difficult to identify and discuss some of the issues in a general manner, and to some extent, share “good practices” or “good systems.” DDD believes that reviewing what individual states do or do not do in terms of their responsibilities and decisions causes concern and heightens the awareness of disparity. We feel this holds equally true for those in special education. The issues, responsibilities, and decisions encountered have their own challenges and vary from state to state. These concerns and disparities also exist for students requiring special instruction.

Clearly and simply, the intent of high stakes testing for all students is to raise expectations for student achievement (i.e., to have all students achieve educational standards set by states and districts) and to monitor student and school progress in order to help raise student achievement on arbitrary standards. In May 2005, Secretary of Education Margaret Spellings announced “A New Commonsense Approach to Raising Achievement for Students with Disabilities” directed toward 2% of school population (in addition to the 1% of students with the most significant cognitive disabilities) who are working toward but may not be able to “reach grade-level achievement standards” (U.S. Department of Education, May 10, 2005). This gives to the states who have applied the flexibility to modify achievement standards and assessments for 2% of their student population. Different responsibilities and decisions that each individual state must make come with this added flexibility.

The federal regulations governing the inclusion of students with disabilities in the NCLB mandated assessments have prompted reconsideration of the role of assessment in special education. Some of these considerations related to assessment have included developing alternate standards and/or modified standards, benchmarks and/or performance indicators, extending special educators’ skills and practices to accommodate the teaching of standards-based content, ensuring that students have access to general curriculum, providing appropriate test accommodations for statewide and school district assessments, and designing and implementing alternate assessments directed toward students with the most
significant cognitive disabilities and “modified” assessments toward students with disabilities who may not be able to achieve grade-level standards.

**Issues and Concerns about Statewide Alternate Assessments**

It is important to recognize that having high expectations of achievement for students with the most significant cognitive disabilities and involving students in statewide assessments that are meaningful (i.e., alternate assessments) is highly supported by educators, parents and associations like the Council for Exceptional Children (CEC) and our Division (DDD) (CEC, 2005; CEC, 2004; Ysseldyke et al., 2004). In devising these state assessments, the U. S. Department of Education (2003) has stated in its regulations that alternate assessments: “should have a clearly defined structure, guidelines for which students may participate, clearly defined scoring criteria and procedures, and a report format that clearly communicates student performance in terms of the academic achievement standards defined by the State. The requirements for high technical quality set forth in §§ 200.2(b) and 200.3(a)(1), including validity, reliability, accessibility, objectivity, and consistency with nationally recognized professional and technical standards, apply to alternate assessments as well as to regular State assessments.” (p. 68699) However, there are various concerns about state alternate assessments and alternate achievement standards that have been developed, or are currently being developed or redefined (e.g., Browder, Fallin, Davis, & Karvonen, 2003; Browder, Wakeman, & Flowers, 2006; Crawford & Tindal, 2006; Kohl, McLaughlin, & Nagle, 2006; Yovanoff & Tindal, 2007). Some of the issues related to state alternate assessments include how they are developed, administered, scored, and reported, and whether they are useful to teachers for improving instruction. Some areas that alternate achievement standards are being questioned on are the validity of their alignment with content standards, and their application to life skills curriculum.

As DDD reviewed the literature, listened to teachers who are responsible for assessing students using their state’s alternate assessment, and discussed critical issues regarding alternate assessments, a number of themes emerge. A major one is the factors that can affect student scores. Browder, Fallin, et al. (2003) identified variables that they felt may influence alternate assessment outcome scores. One of these variables included the technical quality of the alternate assessment format. Yovanoff and Tindal (2007) have agreed stating that “very few research studies have been published on the technical adequacy of alternate assessments…” (p. 186). According to Browder et al. about half of the states have chosen a portfolio format for their alternate assessment. Other states use checklists, IEP analysis and other types of performance-based assessments. More recently, Browder et al. (as cited in Flowers et al., 2005) have identified the most frequently used alternate assessment approaches to be portfolio, performance-based and checklist. A portfolio alternate assessment is a collection of student work gathered to demonstrate the student’s performance on specific skills and knowledge, generally linked to state content standards. Performance-based assessments are items or tasks administered by a teacher or test administrator to a student. These items/tasks are the same for all students. Checklist assessments are a list of skills that an individual familiar with the student’s performance rates according to a specific proficiency scale (Flowers et al.). Yovanoff and Tindal have identified portfolios, observations (teachers are asked to select a behavior that represents student need and then observe it in a functional environment) and performance assessments as the responsive formats used for alternate assessments (see Table 1 for descriptive examples of state alternate assessments). Each of these types of formats has a subjective component and thus we raise a number of questions about assessment considerations. Issues such as rater reliability, the content being assessed and its relation [alignment] to state standards, construct validity, measurement error, assessors’ training, length of assessment, and procedures for administering, scoring and interpreting results must all be addressed and analyzed. It appears that all types of commonly used alternate assessment approaches have problems with technical aspects. Tindal et al. (2003) have been critical of portfolio assess-
# TABLE 1

Descriptive Examples of State Alternate Assessments for Students with Most Significant Cognitive Disabilities

<table>
<thead>
<tr>
<th>State</th>
<th>Portfolio/ Collection of Work</th>
<th>Performance-based Assessment Description</th>
<th>Alternate Assessment Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arizona</strong></td>
<td><em>Portfolio (optional):</em></td>
<td>Data forms are provided for the assessment of student performance. They include: Teacher Evaluation of Student Performance; Activity-Based Performance Assessment (Level I contains reading, writing, listening and speaking and mathematics skills that are infused into functional activities; listening and speaking skills are not assessed for Level II); and Parent Interview.</td>
<td>AIMS-A: Level I—students attending 2nd–12th grades working on Level 1 Functional and Kindergarten level assessments; Level II—students attending 6th–12th grades [whose performance on the Level I assessments meets or exceeds the performance standards] working on articulated standards at the 1st–3rd grades [at a beginning level]. Level I includes a teacher evaluation of student performance on selected standards, a parent interview, and, an evaluation of student performance on four activities in four domains—school or vocational (daily schedule), recreation and leisure (interactive game), community (making purchases at a fast-food restaurant or school snack bar), and domestic living (entertaining friends). Science standards and assessments are currently under development.</td>
</tr>
<tr>
<td><strong>Illinois</strong></td>
<td><em>Portfolio of Student Work</em></td>
<td>Alternate performance indicators (APIs) are listed for each state goal. A team is identified to select the specific APIs that will be focused on for assessment and instruction.</td>
<td>Illinois Alternate Assessment consists of a portfolio of student work and other materials with data collection a minimum of three times during the year. Data must be collected at three points (for baseline, midpoint and ending point) in the school year (with defined dates starting in September and ending early February). The materials can include samples of student work, photos of the student doing work in school or at home and teachers’ summaries of what students have learned. Each portfolio is evaluated on several dimensions:</td>
</tr>
<tr>
<td>State</td>
<td>Portfolio/Collection of Work</td>
<td>Performance-based</td>
<td>Alternate Assessment Description</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Kansas</td>
<td>Data Folio Three pieces of evidence (with at least five trial opportunities for the student to demonstrate the skill) are required for each of the five selected indicators for reading, mathematics and writing alternate assessments.</td>
<td>Indicators are selected from each of the state’s extended standards, two in reading, four in math, and three in writing. The other indicator(s) may be selected from any standard in that curricular area. The IEP team chooses five extended indicators for each of the content areas: reading, mathematics and writing.</td>
<td>The Kansas Alternate Assessment consists of a collection of data (i.e., data folio) of an individual student’s performance of skills and content outlined in the Kansas Extended Standards. Documents are compiled, assembled, scored, and submitted to the state during the spring assessment window. Instruction needs to occur before the assessment window opens and data collection begins. Data for the data folio are collected during the testing window designated by the state department. A scoring rubric (skill level) is used and scored locally by three scorers. Data folio consists of three pieces of evidence for each selected indicator. Formats submitted may include work samples, worksheet, videotape/DVD, photographs, audiotape and data sheets. These products document the student’s knowledge of concepts and content and the performance of skills outlined in the Kansas General and Extended Curricular Standards.</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>No Forty (40) on demand performance tasks per year. Approximately 20 items are related to literacy standards and approximately 20 items are related to numeracy standards. Each test item represents an authentic, relevant, and age-appropriate activity related to reading or mathematics. The science alternate assessment is currently being developed.</td>
<td>The Pennsylvania Alternate System of Assessment (PASA) measures the attainment of knowledge and skills of students with significant cognitive disabilities through a series of approximately 40 discrete items. The PASA is designed to take a snapshot of students’ typical performance on a small sample of academic skills in the areas of reading and mathematics (and soon science) derived from the PA standards. In order for all students with significant disabilities to participate, these tasks are designed with three levels of difficulty to assess skills that are useful in their daily lives. Teachers either scribe or videotape their student’s participation in the on demand performance assessment administered by a test administrator (usually the teacher), one-to-one.</td>
<td></td>
</tr>
</tbody>
</table>
ments in particular because this type of assessment is “difficult to implement with any level of technical adequacy” (p. 483). Similarly, Flowers et al. identified checklists as relying on subjective judgments about student performance instead of direct observation. Also, after working with teachers who conduct performance-based assessments on pre-determined tasks, they have stated that the teachers learn which task levels to implement so that their students perform at the proficient level. The teachers have learned how to use prompts and therefore, feel that the alternate assessment is assessing the teachers’ ability to conduct the assessment and not the student’s level of achievement. Zatta and Pullin (2004) reached similar conclusions after reviewing research related to teachers designing and scoring portfolio assessments. Browder et al. (2006) indicated that one of the needs in improving alternate assessments is “how to establish valid and reliable alternate assessment scores” (p. 256).

Another issue of concern has been identi-
fied as data collection. One of the major questions heard initially about alternate assessment was, “How will the assessment and the data collected help inform teachers for instructional purposes?” This has not only been stated by teachers and other educators who are involved in implementing these assessments but also by researchers (Browder, 2001; Browder, Spooner, et al., 2003; Flowers et al., 2005; Kleinert & Kearns, 2001; Kleinert & Thurlow, 2001). DDD, and others in the field, feel that a major outcome of the alternate assessment should be to improve the quality of our programs for students with the most significant cognitive disabilities. Although in some cases this may be happening, in many cases it is not. For states that used, or continue to use, assessments that are given only once and are not linked directly to the student’s IEP and those standards particularly relevant for that student (e.g., pre-determined performance tasks given to all students with the most significant cognitive disabilities in that state) the results have been discouraging. Some states that use a portfolio to collect data for a number of months (e.g., New Hampshire, New York) or over the course of the school year (e.g., Illinois, North Carolina) are likely to see benefits to this on-going assessment process. However, the data collected must be appropriately linked to both the individual student’s IEP goals and the state standards or alternate content standards and the assessment must be technically sound.

Another major issue of alternate assessments is training which is in turn strongly related to other variables. This training on alternate assessments is multi-faceted and takes in a range of individuals from the students themselves (e.g., self-determination assessments as part of the alternate assessment; involvement in the portfolio assessment process) to teams (e.g., the IEP team’s determination of whether a student is required to take the alternate assessment, and informed decision making regarding the IEP goals).

Although alternate assessment training affects many individuals, the teacher most often has the major responsibility for using alternate standards to guide instruction and ensure that students are accessing the general curriculum. Training includes helping teachers with methodologies and other techniques to incorporate alternate assessments and standards in their day-to-day instruction with students with the most significant cognitive disabilities. Teachers are also responsible for conducting and scoring the alternate assessments, interpreting and reporting the results, and consequently, making decisions based on these results and ensuring that students are involved in the assessment process (e.g., Walsburn-Moses, 2003). Teachers and students must be motivated in this process and therefore need to receive the appropriate training and support to bring about successful assessment. For example, based on the research of Kampfer, Hovrath, Kleinert, and Kearns (2001), Kleinert, Green, Hurte, Clayton and Oetinger (2002) noted that the time teachers spent on assessing alternate assessment portfolios was a slight predictor of student scores. “A far more powerful predictor of student scores was the extent to which the alternate assessment was integrated into daily instruction, as well as the extent to which students were actively involved in the construction of their own assessment portfolios” (Kleinert et al., 2002, p. 40). Crawford and Tindal (2006) found that only 30% of the teachers they surveyed indicated that state department “test results were frequently or always useful in guiding instruction” (p. 215). In 2003, Browder, Fallin, et al. stated that unless teachers are trained to make appropriate connections between the alternate assessments and daily instruction, the assessments may contribute to teacher stress due to time constraints.

In two recent studies to assess teachers’ perceptions about alternate assessments, Flowers et al. (2005) and Toweles-Reeves and Kleinert (2006) found that the alternate assessments infringe on teaching time and teachers’ personal time. As well, teachers using the three main assessment approaches, portfolio, performance-based and checklist, felt that there was an increase demand on paperwork. Flowers and her colleagues recommend that teachers be taught ways to streamline their time and how to better organize and manage the paperwork. They also suggest that states “focus on the most parsimonious ways to document progress on state academic content standards” (p. 90). The same should be noted for school districts as some may be requiring additional
Implications and Recommendations

In addressing only some of the many issues related to alternate assessments, DDD’s position is apparent. States need more time, feedback and resources to fully develop alternate assessments that are relevant to students as well as their teachers, parents and IEP teams. Other needs and implications identified by DDD from the research and the discussions of teachers are: ensuring that functional skills and curriculum are part of the alternate standards and assessments; including parents in the assessment process; providing teachers with more resources and skills; providing training to assist both teachers and students to make connections between the assessment and daily instruction; assisting teachers in helping students be more involved in the assessment process (e.g., self-determination, progress monitoring); ensuring that statewide and district assessments are technically sound (e.g., valid and reliable) and actually assess student’s performance (i.e., not the teacher’s or assessment administrator’s skill in conducting the assessment). Tindal et al. (2003) conclude that more development is needed for alternate assessments before the system can effectively serve teachers and students. The foci should be on teacher training for both implementation and decision making, and on assessments that use “brief measures on a range of relevant behaviors” (p. 493). Browder, Fallin, et al. (2003) state that these measures should be “across time to capture student performance” (p. 263) and IEPs should be based on students’ individual needs and the state standards or alternate state standards. More and more research is being conducted on issues and concerns related to alternate assessments. With this research come additional recommendations as states continue to face the challenge of raising expectations and meeting the goal of proficiency for all of their students.

References


Kleinert, H., Green, P., Hurte, M., Clayton, J., & Oetinger, C. (2002). Creating and using mean-