

Running Head: CONSEQUENCES FROM LARGE-SCALE ALTERNATE ASSESSMENTS

Survey Results from an Analysis of the Consequences from Alternate Assessments

Patricia Almond
Gerald Tindal
Josh Geller

University of Oregon

Consequences of Alternate Assessment

The purpose of this chapter is to report on two studies conducted to address the consequences of alternate assessments judged against alternate achievement standards (AA-AAS) as examples for states. The studies refer to the alternate assessments administered to students with significant cognitive disabilities as part of states' assessment and accountability systems during the 2005-06 school year. The steps and procedures followed in conducting the studies along with the instruments and materials can be adopted or adapted by states for use in conducting similar studies at the state level with special education teachers and administrators. In this brief report, we provide a summary of preliminary findings and methods along with references for the measures.

Overview of the Consequences of Assessment

Dr. Tom Haladyna, an expert on measurement and author of numerous articles and books on assessment, served as a technical advisor for the DAATA project. With Dr. Gerald Tindal, he assembled a volume titled *Large-Scale Assessment Programs for All Students: Validity, Technical Adequacy, and Implementation* (2002), focusing on issues related to making large-scale assessments inclusive. Principles for this chapter on consequences of assessment have been informed and guided by numerous personal communications with Haladyna. The content for this chapter drew greatly from these conversations and from guidance of Project DAATA's national technical advisory committee on which Tom served. He reviewed drafts of this chapter (personal communication, January 2, 2006 and September 11, 2006) and provided valuable guidance about construct irrelevant variance and the consequences of assessment and accountability systems for students with disabilities including those with the most significant cognitive disabilities taking alternate assessments.

The question regarding consequences of assessment involves examining consequences of (a) test use, how the testing program will be used, and (b) test score interpretation, what decisions will be made based on the results. The principles for examining consequences of assessments are the same for students with disabilities as for students without disabilities and alternate assessments should be evaluated by the same criteria for validity as the general assessment. It is clear that the authors of Individuals with Disabilities Education Act 1997, reauthorized in 2004, anticipated positive consequences for including students with disabilities in testing including (a) greater access to the general education curriculum, (b) improved instruction, and (c) higher expectations for students with disabilities. Testing, however, can also have negative consequences, a concern expressed by both parents and educators. For example, consequences can involve decisions for promotion (graduation) and retention, salary incentives for educators, or rewards and sanctions for schools. Other consequences can involve shifts in curriculum, motivation, and morale of teachers due to high stakes and the effects on student motivation. The possible effects of testing (both positive and negative) should be examined and explained to policy makers. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999, 13.1) states, "It is the responsibility of those who mandate the use of tests to monitor their impact and to identify and minimize potential negative consequences. Consequences resulting from the uses of the test, both intended and unintended, should also be examined by the test user" (p. 145).

One question underlying the examination of the consequences of assessment is a concern about low scores. It is important to ensure that low scores do not occur because the alternate assessment measures a student's disability rather than their proficiency on the content being assessed or that the alternate assessment fails to test the intended construct because the test content has been oversimplified and under represented.

Studies of Consequences of Alternate Assessment. For most states, alternate assessments came into existence with the turn of the century (2000-01). Their implementation has outpaced research on their impact and consequences and most of what we do know has come from surveys of educators. What has been reported to date has addressed changes in curriculum and instruction, improvement in student learning, teacher motivation, stress and beliefs about testing, and communication with parents.

Three studies regarding the consequences of alternate assessment employed teacher surveys for learning more about the consequences of alternate assessment. Even though each study was unique, there were some commonalities in the questions asked and the results. The principal question in all three studies centered on the educational benefit to students in taking the alternate assessment and being included in the educational accountability system. Teacher frustration about the amount of time required to conduct portfolio alternate assessments was reported in two studies.

Kleinert, Kennedy, and Kearns (1999) conducted a statewide survey of teachers involved in the Kentucky alternate assessment, the nation's first statewide alternate assessment. Their principal concern was to determine the extent to which the teachers perceived benefits for including their students in state and school accountability measures and their perceptions of the instructional impact of the alternate assessment system upon student outcomes. Results indicated that teachers did realize benefits and saw positive changes in both instructional programming and enhanced student outcomes. Although they did see benefits, teachers expressed frustration with the amount of time required to complete student assessment portfolios and concern about scoring reliability and the extent to which the alternate assessment was more of a teacher assessment than a student assessment.

Roach, Elliott, and Berndt (in press) sought to increase understanding about how teachers perceive and use the Wisconsin Alternate Assessment (WAA). The WAA for the study was a rating scale [consisting of 131 Likert-scale in addition to work samples and observations for each IEP-aligned item. Results are organized into five scales: Reading, Language Arts/Writing, Math, Social Studies, and Science. The teacher survey was conducted as a way to obtain important validity evidence about the consequences of administering the assessment and the instructional utility of the alternate assessment process. In general, teachers in this investigation reported less positive perceptions about the WAA process as students advanced through the grades, and grade 10 teachers seemed to have more reservations about the meaningfulness of WAA scores and their utility for instructional planning. The analysis suggested that teachers confronted three challenges in implementing the alternate assessment: (1) fitting administration into their workload, (2) managing the time demands of the assessment, and (3) ensuring that results reflected student knowledge and skills. Respondents were neutral to somewhat positive about the usefulness of the alternate assessment process for setting instructional priorities.

Flowers, Ahlgrim-Delzell, Browder, and Spooner (2005) also examined teachers' perceptions of alternate assessments. Teachers in five states participated in the teacher survey (CO, KY, NC, NE, and TN). Three states used a portfolio approach as their AA system, one had a performance-based assessment, and one employed a checklist approach.

The results suggest teachers tended to agree that their students should be included in school accountability, but they did not agree with most items assessing the educational benefits of alternate assessments. Teachers reported that the most significant impact of alternate assessments was the increase in paperwork and demands on their time. In contrast, when teachers perceive alternate assessments counting in school accountability systems, more teachers report a positive impact for alternate assessments. Significant differences in the positive impact of alternate assessment between testing approaches (i.e., portfolio, performance-based, checklist) were also found. Implications and recommendations are discussed. Although approximately half of the respondents agreed with including students in accountability systems, few thought that inclusion resulted in higher quality education. As reported in other research, teachers stated that administering alternate assessment (AA) competed with teaching time and individual student needs and that there was an increased paperwork burden, especially for portfolio assessments. Teachers' responses suggest AA has not achieved the benefits for students, teachers, and schools that were desired when IDEA (1997) was introduced. There were differences found among the AA testing approaches and the teachers' perceptions of AA.

Rationale for Research on Consequences of Alternate Assessment

Results of assessment and accountability systems are intended to impact a myriad of education outcomes including the following:

- implemented curriculum
- instructional content and strategies
- classroom assessments
- student, teacher, and administrator motivation and effort
- learning experienced by students
- professional development
- teacher participation in assessment administration and development
- community awareness and beliefs about assessment and the use of test results.

In the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), the consequences of assessment are considered both directly relevant to test validity and important in informing decisions about social policy. Studying consequential evidence informs the value implications of test scores as a basis for setting policy, overseeing the social consequences, as well as the validity of test scores.

Five classes of consequences have been identified: (1) curricular, (2) instructional and IEP reforms, (3) improvement in student learning, (4) teacher motivation, morale, stress, and beliefs about testing process and ethical behavior, and (5) student motivation and communication with parents, constituents, and the public. In addition to documenting consequences, it is important to act on the information collected and use the findings to improve and validate the entire assessment and accountability system. Forte Fast and Hebbler (2004) suggested that in evaluating consequences, questions should, at a minimum, determine the degree to which the

system (a) builds capacity of staff, (b) affects resource allocation, (c) supports high-quality instruction, (d) promotes student equity access to education, (e) minimizes corruption, (f) affects teacher quality, recruitment, and retention, and (g) produces unanticipated outcomes. Baker, Linn, Herman, and Koretz (2002) recommend that longitudinal studies be planned.

Summary of the Studies of Consequences of Assessment

The DAATA Project conducted two studies addressing the consequences of assessment. Both were focused on the question of the impact of alternate assessments in relation to improved instruction and student learning. The first study was a focus group involving members of the Assessing Special Education Students SCASS (State Collaborative on Assessing Student Standards). Members included staff from state agencies and representatives from national organizations and centers. The second was a teacher survey regarding the administration of alternate assessments during the 2005-06 school year. These two studies are summarized below.

Study One—Focus Group

In winter 2006, focus groups were convened to document the perspectives of state representatives and partners on the consequences of alternate assessments. Each subgroup had approximately 10-12 state representatives with the meeting lasting about 90 minutes. Two of the topic areas discussed within one of the focus groups are reported here. The session was moderated using the focus group protocol from Vaughn, Schumm, and Sinagub (1996). Highlights are presented with the topical area and then with a summary of the major issues. Analysis is still underway on the focus group results. Two additional topic areas (focus on technical adequacy and looking to the future) were addressed and will be reported upon completion of the analysis on these areas and a synthesis of themes across subgroups.

Topical Area—Alternate Assessment (AA) in Relation to General Assessment: We would like to get a sense of what each of you think about the design of your state's alternate assessment and its relation to the general education assessment. One way of looking at an AA is to think of it as an extension of the general education assessment. Another way to think of the alternate assessment is as a test that involves a balance between functional skills and academic content. How do you see your state's 2005-2006 alternate assessment?

Clearly, there has been more emphasis on academic outcomes with considerable shift, moving from functional to more academic contexts as the central assessment focus. States are beginning to screen more carefully the assessment tasks being used as representations of academic contexts. No longer is a simple sorting task (e.g., distinguishing among eating utensils) considered as a representation of mathematics). Indeed, the opposite outcome appears to be the case: The academic focus is now central and the functional aspects need to fit within the academic focus.

Quite consistently, the states discussed the breadth and depth of standards being assessed and the way in which the alternate assessment was being mapped to the standards. Several states noted science was being developed in the general education assessment and therefore was being assessed in the alternate assessment for student with the most significant cognitive disabilities. This parallelism of the grade level standards, grade level tests, and the alternate assessment was emphasized and endorsed by virtually everyone in the group.

Considerable discussion was held on the use of “essence” or “extended” standards. These terms were both considered as being instrumental to accessing the general education content standards and considered critical to guide instruction. No consistent definitions were achieved on these two terms, and both agreement and disagreement emerged about the use of the terms. On the one hand, it appeared that the focus of the alternate assessment was NOT the same as the “written” standard; yet, there was general agreement that the alternate assessment reflected the heart of the standards. In this process, it might become apparent that the standards are not addressed with the same depth or breadth in the alternate assessment as in the general education assessment. The term “essence” was used but there was general agreement about the need for a common language.

Topical Area—Student Learning: One intention of recent federal legislation was to bring about improved student learning in academic content for students with significant cognitive disabilities, particularly reading, writing, and mathematics. In what ways do you think that development and administration of the alternate assessment has positively influenced student learning in academic content until now, the current year 2005-2006?

Respondents were quite certain and consistent in stating that in the last year, a substantial focus has been on academic learning standards and goals and even more specifically, in the areas of reading and mathematics. States expressed considerable support for student improvement in reading and mathematics. In addition, they noted that students are no longer learning simply access skills but are learning the academic standards as well.

Focus group respondents in this discussion were state agency representatives. They reported that the academic focus has resulted in professional development that was clearly oriented to the teacher level (perhaps more than at the student level). Questions arose, however, about the curriculum (often there was none). With a focus on instruction comes an emphasis on professional development and the need to integrate this new focus on academic content into Individualized Educational Programs (IEPs). Comments also addressed the connection between special and general education teachers and programs. This topic also generated interest by state personnel in focusing on networks within the state to both gain traction and ensure a common message as well as common language. Some of this professional development may invite teachers to review the alternate assessments and reflect upon the presence of the academic standards in reading and mathematics.

Nevertheless, there still may be a gap between the immediate achievement and eventual achievement. When student examples are collected, they can be used formatively and provide the impetus of next generation changes. However, some of the changes in learning may reflect teachers teaching to the test so the question remains whether the outcomes are real and durable. Another challenge raised was the context in which the alternate assessments are enacted and the need for states to maintain an academic focus through iterations of their support materials (e.g., each year, providing the field with versions that are more clear and focused).

Study Two—Teacher Survey:

The survey was conducted in April and May 2006 to investigate the effect of administering alternate assessments to students with significant cognitive disabilities, particularly as it relates

to instruction and to learning content in reading, writing, and mathematics. The survey addressed the impact of alternate assessment on instruction and learning in four areas: (a) teacher *knowledge and instructional practice* related to academic content in reading, writing, mathematics, and science for students with significant cognitive disabilities, (b) teacher *experience and perception* about administering the current state alternate assessment to students with significant cognitive disabilities, (c) teacher *familiarity with the statewide assessment system* (both the general assessment and the alternate assessment), and (d) *areas of professional development* that would support teachers in their instruction and assessment in academic content for students with significant cognitive disabilities.

The target population was special education teachers who administered at least one alternate assessment during 2005-06. The teacher survey was conducted across eight states (AK, OH, NM, RI, SC, SD, WA, WY), and 976 special education teachers responded. Many respondents (42%) reported having 11 or more years and one-third had five or fewer years of experience with students with significant cognitive disabilities. More than half of the respondents held master's degrees and 94% were white. Approximately 60% of the respondents were not confident or only somewhat confident in explaining assessment results to parents. When asked whether they saw benefits in having their students included in the accountability process for their school, nearly half (47%) did not see student benefits. More than half of the teachers indicated their student case load was between 6 and 12 students and approximately one-half taught students in the elementary grades with the other half teaching in middle or high school.

Teacher respondents were asked how much time their students with significant cognitive disabilities spent on academics (reading, writing, mathematics, science, and/or social studies) per week. Table 1 below shows their responses. There were 47 respondents who did not answer this question. More than 60% who responded indicated that students spent five or more hours per week on academic content.

Table 1. Time Spent on Academics

| Hours/Week | Frequency | Percent |
|-----------------|-----------|---------|
| None | 24 | 2.6 |
| 1-2 Hours | 125 | 13.4 |
| 3-4 Hours | 188 | 20.2 |
| 5-6 Hours | 207 | 22.3 |
| 7 or More Hours | 386 | 41.5 |
| Total | 930 | 100 |

Respondents were asked about changes to their teaching and instructional practices following the passage of the No Child Left Behind Act and the IDEA 1997 requirements for access to the general education curriculum and administration of alternate assessments. Some respondents (approximately 10 %) were first-year teachers and were unable to reflect on changes in their teaching and instructional practices.

Table 2. Changes to Teaching Practice (N = 912)

| Statement | Disagree | Agree | New Teacher |
|---|----------|-------|-------------|
| I spend more time preparing and teaching academic content to students with significant cognitive disabilities. | 29.7% | 60.2% | 10.1% |
| I provide more supports during instruction and test administration. | 28.6% | 61.4% | 10.0% |
| I have higher expectations about what students with significant cognitive disabilities know and can do in academic content. | 37.0% | 53.1% | 9.9% |

These findings demonstrate two approaches to gathering evidence about the consequences resulting from the administration of alternate assessments with a focus on student learning and academic content. Although many teachers who gave alternate assessments indicated that their students spend five or more hours per week on academic content and that their instructional practices have shifted to reflect a focus on student learning in academic content, a lower percent reported “higher expectations” for students with significant cognitive disabilities. In their open-ended responses to two survey questions, many teachers reflected on their sense of benefits for students focusing on two areas in particular. First, respondents expressed reservations about the relevance of division in arithmetic and main ideas in reading for students who are pre-symbolic and are 18 years of age, noting that these students will soon leave school and will need skills to help them succeed in post secondary environments. One respondent did reflect on the “standards based” focus of the alternate assessment:

I am very impressed with my state's approach to alternative [alternate] assessment because it incorporates academic achievement with functional daily living practices. I am very happy that my state has expanded performance standards so that everything we do as a classroom can be and is standards based just like the rest of the school curriculum.

Issues raised by earlier studies were reiterated. Alternate assessments, particularly portfolios, are time consuming and, according to some teachers, take time and effort away from instruction. Teachers often had the sense that the assessments measured the teachers more than the students and that the paperwork involved increased a teacher’s burden without a clear benefit to the students. The findings continue to support the need for professional development about standards based alternate assessments, academic emphasis that “makes sense” to teachers, and the need for more efficient administration methods that result in data about student achievement.

Recommendations for Collecting Evidence about the Consequences of Assessment Programs

It is reasonable to evaluate the extent to which empirical evidence suggests positive and/or negative consequences resulting from administration of the AA-AAS. In doing so, it is important to review documents describing the purpose and intended outcomes of the assessment program. Certainly, improving instruction and student achievement are the ultimate goals of an accountability system, however, an evaluation of the curriculum, standards, or learning outcomes allows an evaluator to reflect on the specific goals of the educational program. In addition,

evaluating professional development, resources, and support places the evaluation in the appropriate context. Much of the consequential evidence is based on surveys of teachers and principals. Surveys are cost effective and increase generalizability because of their typically large sample sizes. Caution is warranted because surveys can only capture the perceptions of the respondents and may not reflect the actual effects of the alternate assessment. More direct findings in evaluating consequences can be found in evidence obtained through the analyses of classroom instruction and assessment activities. In addition, case studies have the potential of capturing the complexity of instructional practices and can illustrate factors contributing to quality instruction and student learning.

References

- American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.
- Baker, E. L., Linn, R. L., Herman, J. L., & Koretz, D. (2002). *Standards for educational accountability systems* [Policy Brief 5]. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.
- Flowers, C., Ahlgrim-Dezell, L., Browder, D., & Spooner, F. (2005). Teachers' perceptions of alternate assessments. *Research and Practice for Persons with Severe Disabilities*, 30(2), 81-92.
- Forte Fast, E., & Hebbler, S. (2004). *A framework for examining validity in state accountability systems*. Washington, DC: Council of Chief State School Officers.
- Kleinert, H. L., Kennedy, S., & Kearns, J. F. (1999). The impact of alternate assessments: A statewide teacher survey. *Journal of Special Education*, 33, 93-102.
- Roach, A.T., Elliott, S.N., & Berndt, S.A. (in press). Teacher perceptions and the consequential validity of an alternate assessment for students with significant disabilities. *Journal of Disability Policy Studies*.
- Tindal, G., & Haladyna, T. (2002). *Large-scale assessment programs for all students: Validity, technical adequacy and implementation*. Mahwah, NJ: Lawrence Erlbaum.
- Vaughn, S., Shay Schumm, J., & Sinagub, J. (1996). *Focus group interviews in education and psychology*. Thousand Oaks, CA: Sage.