In conjunction with the Individuals with Disabilities Education Act of 1997 (IDEA), the No Child Left Behind Act of 2002 (NCLB) calls for specific enhancements to existing assessment and accountability systems within definite timelines. The NCLB Act also stresses the use of reliable and valid data for decision making. The purpose of this study was to examine the consequential validity of the alternate assessment system in one Midwestern state. Three hundred and four teachers completed a survey designed to measure their perceptions of the alternate assessment’s influence on instruction and to see where future training might need to focus in order to improve instruction and alternate assessment results. Findings indicate the alternate assessment influenced instruction to a greater degree than Individualized Education Plan development. A longitudinal analysis of the scoring distribution for alternate assessment results from the 1997-2002 school years was also conducted. Findings and implications are discussed. An appendix contains the teacher survey. (Contains 2 figures and 39 references.) (SLD)
What Are the Consequences? Validation of Large-Scale Alternate Assessment Systems and Their Influence on Instruction

Elizabeth A. Towles
Brent Garrett
Paula Burdette
Mike Burdge
University of Kentucky
Abstract

In conjunction with the Individuals with Disabilities Education Act of 1997 (IDEA), the No Child Left Behind Act of 2002 (NCLB) calls for specific enhancements to existing assessment and accountability systems within definite timelines. The NCLB Act also stresses the use of reliable and valid data for decision-making. The purpose of this paper is to examine the consequential validity of the alternate assessment system in one Midwestern state. Three hundred and four teachers completed a survey designed to measure their perceptions of the alternate assessment’s influence on instruction and to see where future training might need to focus in order to improve instruction and alternate assessment results. Findings indicated the alternate assessment influenced instruction to a greater degree than IEP development. A longitudinal analysis of the scoring distribution for alternate assessment results from the 1997-2002 school years was also conducted. Findings and implications are discussed.
What Are the Consequences? Validation of Large-Scale Alternate Assessment Systems and Their Influence on Instruction

Given the scope of the reauthorization of the Elementary and Secondary Education Act of 1994 (ESEA), also known as the No Child Left Behind Act of 2002 (NCLB), it seems an appropriate beginning to a discussion of validity issues of large-scale assessment systems, in particular, alternate assessment systems for students with significant cognitive disabilities. In conjunction with the Individuals with Disabilities Education Act of 1997 (IDEA), the NCLB Act calls for specific enhancements to existing assessment and accountability systems within definite timelines. The reforms brought into motion by NCLB are structured around four principles of increased accountability, flexibility for states and local schools, expanded options for parents, and an emphasis on data supported methodology (Figlio, 2002).

The term large-scale assessment system generally refers to one manner in which educational accountability is measured. In most state large-scale assessment systems, there are three participation options for students with disabilities: (1) general assessment without accommodations, (2) general assessment with accommodations, or (3) alternate assessment (Thompson & Thurlow, 2001). Students participating in most states’ alternate assessment systems have significant cognitive disabilities. However, alternate assessment systems vary across states creating a need for studies to examine different state systems, particularly with regard to technical issues such as reliability and validity.

The NCLB Act stresses the use of reliable and valid data for decision-making. Scores must be valid for individual student reporting, as well as reporting of aggregated scores by student groups based on poverty, race and ethnicity, disability, and limited
English proficiency. NCLB also requires that state assessment systems, including alternate assessments, “be valid for the purposes for which the assessment system is used; be consistent with relevant, nationally recognized professional and technical standards, and be supported by evidence . . . of adequate technical quality for each purpose” (NCLB, 2002 §200.2(b)(4)(i,ii)). If alternate assessment results are to be included in overall accountability systems, the results must be reliable and valid for the purposes for which they are intended. Considering that the trend to connect assessment systems, including alternate assessments, to high stakes for students, the consequences of how these assessments are scored and the consequences of how these scores are utilized is of utmost importance to students with disabilities unable to participate in the general assessment system, even with accommodations (Quenemoen, Rigney, & Thurlow, 2002).

Research into alternate assessments has focused primarily on implementing alternate assessments (Kleinert & Kearns, 2001; Thompson & Quenemoen, 2001) and the impact of assessment on instruction (Kampfer, Horvath, Kleinert, & Kearns, 2000; Kleinert & Kearns, 2001, 1999; Kleinert, Kennedy, & Kearns, 1999). Research on the reliability of alternate assessments is minimal, but developing (Crossen, Garrett, Hurte, & Peters, 2002; Garrett, Towles, Kleinert, & Kearns, 2003). Some research is beginning to emerge on the scoring and reporting of assessment results (Bechtard, 2001; Burdette, 2001). However, a thorough analysis of the validity of alternate assessments is noticeably absent from the assessment and validity literature.

The only other study we are aware of that addressed the consequential validity of alternate assessments was by Kleinert et al. (2001). Results of the study indicated that teachers from schools that had high alternate assessment scores and high overall school
accountability perceived the alternate assessment to have a positive impact on instruction, collaboration, and the inclusion of students with disabilities. As might be expected, teachers from low performing schools saw the alternate assessment as having little effect in their schools. Principals had mixed perceptions, but those from high performing schools found the assessment to have a more positive influence than principals from low performing schools.

For the purpose of this study, we define validity as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (AERA, APA, & NCME, 1999, p. 9). Stated simply, we will explore what stakeholders in numerous states perceive to be the intended and unintended consequences of the assessment – or the consequential validity of the assessment. While there is some controversy over the use of perception data to measure validity (Linn, 1997; Mehrens, 1997; Popham, 1997; Shepard, 1997), there appears to be overwhelming support for the importance of addressing what we call the “so what” question – did the assessment achieve the purpose for which it was intended (Haertel, 1999; Kane, 2002; Kleinert et al. 2001; Lane, Parke, & Stone, 1998; Lane & Stone, 2002; Linn, 1998; Linn, 1997; Ryan, 2002; Shepard, 1997)?

NCLB also requires that state assessment systems, including alternate assessments, “be valid for the purposes for which the assessment system is used, be consistent with relevant, nationally recognized professional and technical standards, and be supported by evidence . . . of adequate technical quality for each purpose” (NCLB, 2002 §200.2(b)(4)(i,ii)). If state assessment systems, including the alternate assessment portfolios, are valid for the purposes for which they are intended, then consequential
validity becomes an important aspect of the validity discussion concerning large-scale alternate assessment systems.

Large-scale assessments have become a tool used by state and federal policy makers to motivate education reform, as well as a strategy for increased educational accountability (Haertel, 1999; Linn, 1993). Assessments ultimately affect what teachers teach and what students learn (Moss, 1992). Therefore, in order to influence instructional practices at the school level, education reformers often use large-scale assessments to link systematic education reform and classroom instruction.

Purpose of the Paper

This exploratory study examines the consequential validity of the alternate assessment. Of particular interest is to what degree were teachers’ instructional practices and IEP development influenced by the alternate assessment system. Kane (2002) suggests that what is most important in developing large-scale assessment systems, particularly high-stakes systems, is generating a sense of “legitimacy of using the actual or anticipated consequences of the testing program” to make policy decisions. In the current study, we examined one Midwestern state that developed its alternate assessment system in order to provide greater levels of educational and social opportunities for students with significant cognitive disabilities (Kleinert, Kearns, & Kennedy, 1997) and to determine if the consequences of that assessment system were, in fact, improved educational opportunities. Using teacher surveys and assessment outcome data, we analyzed the validity of the alternate assessment system in that state to determine the influence of the assessment on teacher instruction.

Methods

BEST COPY AVAILABLE
To determine the influence of large-scale alternate assessment systems on instruction received by students with significant cognitive disabilities, teachers having one or more students eligible for the alternate assessment in the 2001-2002 school year in one Midwestern state were administered a survey.

Sample

The sample for this study included elementary, middle, and high school teachers who participated in four regional scoring sessions of the 2002 alternate assessment in this state. All respondents were special education teachers, primarily teaching students with significant cognitive disabilities. Three hundred and four teachers completed the survey, providing an almost 100% response rate (304/307).

Instrument

Staff from this particular state’s alternate assessment system developed the survey instrument (refer to Appendix A to review the survey). The purpose of the survey was to measure teachers’ perceptions of the alternate assessment’s influence on their instruction and to see where future training might need to focus in order to improve instruction and alternate assessment results. Survey questions were developed using three procedures:

- Investigating the alternate assessment process overall in terms of both daily instruction and IEP development,
- Investigating each aspect of the alternate assessment scoring rubric (dimensions such as standards, performance, settings, support, social relationships, and self determination) in terms of both daily instruction and IEP development, and
• Reflecting upon possible reasons for low ratings of specific rubric dimensions (i.e., don’t know how to implement, don’t have the needed support to implement, don’t think it’s important).

Data Analysis

We computed and graphed mean statistics to measure the degree to which the assessment and each dimension of the assessment was correlated with IEP development and instruction by special education teachers with students participating in the alternate assessment system. In order to validate the survey findings, we also conducted a longitudinal analysis of the scoring distribution for alternate assessment results from the 1997-2002 school years.

Results

Figure 1 presents the results of the teacher survey. Findings indicated the alternate assessment influenced instruction to a greater degree than IEP development. The six dimensions of the alternate assessment varied in their influence on instruction. For each dimension, as for the assessment as a whole, teachers rated instruction as more influenced by the assessment than IEP development. The performance (M=3.86, M=3.70), self-determination (M=3.78, M=3.59), settings (M=3.79, M=3.61), and standards (M=3.78, M=3.65) dimensions were perceived to have the most influence on instruction and IEP development respectively. The support (M=3.72, M=3.52) and social relationship (M=3.67, M=3.48) dimensions were perceived to have less influence on both instruction and IEP development.

The alternate assessment system scoring distribution for 1997 through 2002 is graphed in Figure 2. The figure highlights two trends. First, the percentage of apprentice...
assessments dropped considerably over the six-year time span (51% in 1997 to 25% in 2002). Concurrently, the percentage of distinguished assessments increased dramatically (8% in 1997 to 28% in 2002). There has been little change in the percentage of proficient assessments (21% in 1997 to 26% in 2002) while novice scores slightly increased (14% in 1997 to 21% in 2002).

Discussion

Looking at both the intended and unintended consequences of implementing alternate assessment systems may begin to focus the true effects of NCLB at both the teacher and student level. Teachers perceived alternate assessments to have a stronger influence on instruction than IEP development. This finding bears evidence of the link between the intended purpose of the alternate assessment, which is to provide increased educational opportunities for students with disabilities who are not able to participate in the general assessment even with accommodations, and changes to instruction within the classroom.

It is interesting that alternate assessments were perceived to have a lesser degree of influence on IEP development. One reason for this finding may be the lack of a clear link between the two. Some states link the alternate assessment to the IEP while others use checklists or portfolios (Thompson & Thurlow, 2001). Browder, Fallin, Davis, and Karvonen (in press) clearly express “it may be difficult to find a link between the alternate assessment and the record of progress teachers have obtained while tracking IEP objectives. If the alternate assessment does not have a clear link to the IEP, teachers may need guidelines on how to track progress for this new requirement” (in press). As a result,
one aspect of future teacher training may need to include information on how to track progress through IEP objectives in a reliable and valid manner.

Another caveat to this discussion involves the hiring of teachers out of field or through emergency certificates. Due to teacher shortages around the nation, many students, more specifically those with significant cognitive disabilities, may not be receiving the most effective instructional strategies recently outlined by numerous educators (e.g., Cipani & Spooner, 1994; Snell & Brown, 2000; Westling & Fox, 2000). MIKE ARE THERE ANY OTHERS WE CAN CITE HERE? Even those teachers who are committed to sound educational practices “may be confused about the link between instruction and alternate assessment outcomes especially when there is no defined relationship to the IEP in their states’ process” (Browder et al., in press, p. 13). Together, linking the IEP to assessment may be difficult for all teachers and perceived as less influential when compared with instruction.

Findings from this study also suggested a trend in the alternate assessment score data in this particular state. The number of distinguished portfolios has increased over time while the number of apprentice portfolios has decreased dramatically. Since the number of distinguished portfolio scores within this state has increased in the past six years, ideally, it is hoped teacher instruction has improved and educational opportunities for students with disabilities have increased due to the implementation of the assessment system. However, an unintended consequence to the implementation of the alternate assessment system may be the fact that teachers are beginning to understand exactly how to document the evidence required in performance-based assessments, such as portfolios, and can actually complete the portfolio to make it appear as if educational opportunities
are actually increasing and student outcomes are improved. It may be best to hypothesize that both are taking place.

Limitations

This exploratory study provides important preliminary insight into the consequential validity of alternate assessments. Future studies can improve upon this work by incorporating the views of multiple stakeholders into the study design (Kane, 2002; Lane & Stone, 2002). As this study emerged from conversations with teachers regarding the impact of alternate assessments on student planning and instruction, it was a logical viewpoint with which to begin. Stakeholders such as principals, district- and state-level administrators, parents, and teachers should be included in future analyses.

In addition, a qualitative component needs to be added to the research to explore the unintended consequences of the alternate assessment system (Lane, Parke, & Stone, 1998; Lane & Stone, 2002). A study by Lane and Stone (1998) suggested that unintended consequences are possible, such as (1) the narrowing of curriculum and instruction to focus only on the specific learning outcomes assessed, (2) the use of test preparation materials which are closely linked to the assessment without making changes to the curriculum and instruction, (3) the use of unethical test preparation materials, (4) and inappropriate use of test scores by administrators.

Conclusion

NCLB calls for specific enhancements to assessment systems, in particular alternate assessments, and stresses the use of reliable and valid data for decision-making. The current study was one example of how researchers are trying to better understand the consequential validity of the alternate assessment in one Midwestern state with hopes to
improve upon the system and increase outcomes for students with significant cognitive disabilities. Results suggested assessment had a greater influence on instruction than IEP development. Findings from a longitudinal analysis also indicated trend data in alternate portfolio scores from 1997 through the 2002 school year.
References


Browder, D., Fallen, Davis, & Karvonen (in press)


Langenfeld, K., Thurlow, M., & Scott, D. (1997). High stakes testing for students:
Unanswered questions and implications for students with disabilities (Synthesis Report 26). Minneapolis: University of Minnesota, National Center of Educational Outcomes.


Appendix A

Alternate Assessment Teacher Survey

Please circle the answer that best fits. Please circle only one answer unless otherwise specified.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>No influence</th>
<th></th>
<th>High degree of influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alternate Assessment has influenced my IEPs</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>The Standards Dimension has influenced my IEP</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>The Performance Dimension has influenced my IEP</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>The Support Dimension has influenced my IEP</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>The Social Relationships Dimension has</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>influenced my IEP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>The Self-Determination Dimension has</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>influenced my IEP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Alternate Assessment has influenced my</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>overall instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>The Standards Dimension has influenced my</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>The Performance Dimension has influenced my</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>The Settings Dimension has influenced my</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>The Support Dimension has influenced my</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>The Social Relationship Dimension has</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>influenced my instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>The Self-Determination Dimension has</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>influenced my instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Mean Response Rates to the Degree to Which the Alternate Assessment Influenced IEP Development and Instruction

Assessment Categories

Influenced IEPs
Influenced Instruction
Figure 2: Scoring Distribution for Alternate Assessment System
I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Elizabeth Towles

Organization/Address: University of Kentucky

Lexington, Kentucky 40506

Printed Name/Position/Title: Elizabeth Towles/RA/Res. + Eval. Asst.

Telephone: 606-257-7672

Fax: 606-257-4939

E-Mail Address: etawes1@uky.edu

Date: 03/30/03
III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

<table>
<thead>
<tr>
<th>Publisher/Distributor:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Address:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Price:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

<table>
<thead>
<tr>
<th>Name:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Address:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse: University of Maryland

ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions

088 (Rev. 2/2001)