

2004

An analysis of alternate assessment policy: Findings from six states

Tamra Roberts Cobb
William & Mary - School of Education

Follow this and additional works at: <https://scholarworks.wm.edu/etd>



Part of the [Special Education and Teaching Commons](#)

Recommended Citation

Cobb, Tamra Roberts, "An analysis of alternate assessment policy: Findings from six states" (2004). *Dissertations, Theses, and Masters Projects*. William & Mary. Paper 1550154041. <https://dx.doi.org/doi:10.25774/w4-3b7k-rj98>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.

**AN ANALYSIS OF ALTERNATE ASSESSMENT POLICY:
FINDINGS FROM SIX STATES**

A Dissertation

Presented to

The Faculty of the School of Education
The College of William and Mary in Virginia

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

by

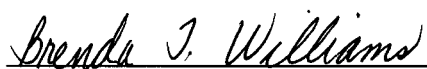
Tamra Roberts Cobb

April, 2004

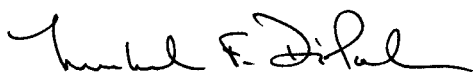
**AN ANALYSIS OF ALTERNATE ASSESSMENT POLICY:
FINDINGS FROM SIX STATES**

By Tamra Roberts Cobb

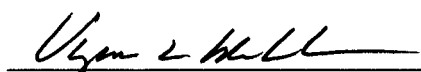
Approved April 8, 2004 by



Brenda T. Williams, Ed.D.
Chair of Dissertation Committee



Michael F. DiPaola, Ed.D.



Virginia L. McLaughlin, Ed.D.

DEDICATION

This dissertation is dedicated to my loved ones, family, and friends for your support and encouragement during this arduous process. And in loving memory of my maternal grandmother, Lucy Coles, who always wanted me to have children. Mama Lucy, this is my baby; to my parents, Elaine and Gentry Roberts, my ultimate supporters and cheerleaders; to my husband, Gary, who put his life on the line to permit me to see this through; to my in-laws, Audrey and Jett Cobb, for your continuous encouragement; to my brother Mark, for dropping off an assortment of forms, drafts, and documents on campus when I could not; to his wife, Cherise, for helping me understand what it is like to participate in the alternate assessment program as a special educator; to my niece and nephew, Brittany and Marquis, for forgiving Aunt Tamra for missing all of your activities and programs; and to my girls - Loury and Lisa - for always being there and sharing this with me.

TABLE OF CONTENTS

Acknowledgments	vi
List of Figures	vii
List of Tables	vii
List of Appendices	viii
Abstract	ix
Chapter I - The Problem	2
Context for Reform in Assessment and Accountability	4
State Implementation of Alternate Assessment	6
Research on Alternate Assessment Development and Implementation	6
Statement of the Purpose	8
Research Questions	9
Significance of the Study	10
Definitions	12
Limitations, Delimitations, and Assumptions	14
Chapter II - Review of the Literature	17
Impact of Special Education Reform on Assessment Systems	17
Legislative Background	19
Policy Implementation	25
Inclusive Accountability Systems	26
Implementation of Alternate Assessment	27
Core Principles and Characteristics of Assessment and Accountability Systems	42
Summary of Literature Review	56
Chapter III - Methods	59
Design	60
Instrumentation and Protocol	62
Data Analysis	73
Ethical Safeguards	78

Chapter IV - Data Analysis	79
Phase One	80
Phase Two	81
Phase Three	84
Summary of Analysis	96
Chapter V - Findings, Recommendations, and Implications	99
Findings	99
Recommendations and Implications for Future Research	103
Recommendations and Implications for Policy	106
Closing Statements	109
Appendices	111
References	129
Vita	147

ACKNOWLEDGMENTS

There are numerous individuals who have helped me stay the course during this journey; far too many to single out. Yet, there are those who cannot go unnamed. Dr. Williams, thank you for encouraging me to pursue the degree when I questioned what an “outsider” could contribute to the development of educational leaders, much less become one. Thank you to the EPPL Special Education cohort and our adopted general education members for sharing your wisdom and friendship and being study and test pilot partners. To Dr. Patton, thank you for your support and mentorship; and especially for the opportunity to co-author an article that was subsequently published. Dr. Finnegan, thank you for making me want to explore everything beneath its surface, not only allowing me to uncover meaning but also enabling me to create my own significance. To Sharon Siler at the Virginia Department of Education, thank you for exposing me to and immersing me in the world of alternate assessment. I have you to thank (or maybe blame!) for my decision to pursue this area of research.

I also thank all of the state and national administrators, leaders, and policymakers who were willing to share their thoughts on future directions, research needs, and the implementation of alternate assessment policy. Special thanks to Rachel Quenemoen and Martha Thurlow, Eileen Ahern, Ken Olson, and Nancy Arnold for responding to my emails, calls, and early pilot documents.

Finally, thank you to my committee, who was there when I needed them to question, probe, critique, and reassure me. To Dr. Brenda Williams, my chair, thank you for reminding me “... there are as many opinions out there as there are scholars.” Thank

you Dr. Michael DiPaola for “just asking,” and Dr. Virginia McLaughlin for being a tough but gentle spirit. I could not have asked for a better committee.

LIST OF FIGURES

Figure 1	Core Themes of Inclusive Assessment and Accountability Systems.....	95
----------	---	----

LIST OF TABLES

Table 1	Literature Addressing the Core Principles.....	55
Table 2	Examples of State Documents Included in the Content Analysis.....	69
Table 3	Specifications for Data Analysis.....	77
Table 4	Core Principles Noted in State Policy Documents.....	82
Table 5	Indicators Ranked in Order of Prevalence by State.....	84
Table 6	Results of Open and Axial Coding.....	88

LIST OF APPENDICES

Appendix A: Guidelines for Coding.....	112
Appendix B: Coding Instructions.....	119
Appendix C: Content Analysis Coding Form.....	121
Appendix D: Documents Included in the Content Analysis.....	123
Appendix E: Summary of Categories, Codes, and Indicators.....	126
Appendix F: Summary of Test Coding with Second Coder.....	128

**AN ANALYSIS OF ALTERNATE ASSESSMENT POLICY: FINDINGS
FROM SIX STATES**

ABSTRACT

Under the No Child Left Behind Act of 2001 (NCLB, 2002) state assessment plans must demonstrate high standards and challenging content for all students. Additionally, states must include and report performance of all students regardless of how they participate. Since its mandated implementation date of July 2000, alternate assessment has been, and will continue to be, used to assess students unable to participate in the general assessment. Much of the research on the status of alternate assessment has been conducted by the National Center on Educational Outcomes, which in turn has developed what it believes to be principles of inclusive assessment and accountability systems. Because NCLB requires that alternate assessment results be included in accountability indices that reflect adequate yearly progress for all schools, states must revisit their programs within the context of emerging policy and research. This study examined state alternate assessment policy using content analysis and grounded theory approaches to determine the existence of principles of inclusive assessment and accountability systems reported in the literature.

TAMRA ROBERTS COBB

SCHOOL OF EDUCATION, EDUCATION POLICY, PLANNING & LEADERSHIP

THE COLLEGE OF WILLIAM AND MARY IN VIRGINIA

**AN ANALYSIS OF ALTERNATE ASSESSMENT POLICY: FINDINGS FROM
SIX STATES**

CHAPTER I - The Problem

The forces shaping educational reform have been steady, but the focus has shifted dramatically over the last 20 years. Throughout the years legislative mandates, federal initiatives, state and local politicians, educators, families, the business community, and advocates have driven decisions and practice in education. Their push has been spurred on by conflicting beliefs about merit versus equality. Over the past 20 years, the focus has changed from process, to context, to accountability. Thus, public education has gone from excluding students with disabilities, to providing these students access to the general curriculum, to ensuring their meaningful participation in assessments, to requiring system accountability for all students.

Concerns about education and student achievement have been voiced for over three decades; however, stakeholders have generally been at odds regarding the underlying problem and have left policymakers to sort it out. While students in general education have typically benefited from existing standards reform efforts, students with disabilities have not (Furney, Hasazi, Clark/Keefe, & Hartnett, 2003). Unfortunately, the concerns of those involved with students with disabilities have been absent in policy-making discussions.

This ongoing tension between the perceived problems and desired solutions and the existence of a dual accountability system has led to the need to address the education and achievement of all students in a single piece of legislation - the No Child Left Behind

Act (NCLB) of 2001 (NCLB, 2002). With the focus on high standards and ways to measure progress toward them came this legislation, which mandates the inclusion of all students in state- and districtwide assessment programs, accountability systems, and annual testing. Thus, for the first time in U.S. history, any state wishing to receive federal Title I dollars must have in place a plan that demonstrates the same challenging content and achievement standards for *all* students. Further, NCLB requires that each state have a single, statewide accountability system.

Making a distinction between assessment programs, accountability systems, and annual testing is necessary for understanding how reform has steered education from process and context to results. “Accountability and assessment are not the same thing, and testing is just one type of assessment” (Thurlow, Elliott, & Ysseldyke, 1998, p. 2). Assessment involves collecting information about students’ abilities in order to make decisions (e.g., regarding eligibility of services) using such means as classroom tests, observation, and published tests (Elliott, Braden, & White, 2001; Thompson & Thurlow, 2003b). Thompson and Thurlow (2003b) further specify that assessments not only include classroom tests and eligibility assessments but also large-scale assessments at the national, state, and district levels.

Accountability, on the other hand, is broader than assessment, and generally includes more than just the information gathered via assessment. Accountability has been described as a systematic means of collecting data, analyzing it, and using it to assure those inside and outside the educational system that schools are advancing in desired directions (Committee for Economic Development [CED], 2001; Elliot et al., 2001; Elliott & Thurlow, 2000; Hill & DePascale, 2003, Thurlow, Elliott et al., 1998). Both

system and student accountability are often addressed in educational reform efforts, but the requirements under NCLB focus primarily on system accountability. In order to produce meaningful learning experiences for all students, they all must be included in assessments and their results reported along with the results of students in general assessment programs (Burgess & Kennedy, 1998; Thompson & Thurlow, 2003b; Thurlow et al., 1996).

Since the implementation of the Individuals with Disabilities Education Act (IDEA) in 1997, states have used the alternate assessment to allow students with the most significant disabilities to participate in state assessment programs. However, these results have generally not been included in state accountability data. This study examined alternate assessment programs in six states; Kentucky, Maryland, North Carolina, Tennessee, Virginia, and West Virginia, to determine the extent to which core principles and characteristics of inclusive assessment and accountability systems reported in the literature are evident in state policies and documentation.

Context for Reform in Assessment and Accountability

Legislative Background

On January 8, 2002, the NCLB Act of 2001 (NCLB, 2002) went into effect calling for increased accountability via annual testing and reports of progress by subgroups, including special education. Further, NCLB required that states track school districts' yearly progress toward achievement goals, thereby heightening the challenges that states face in ensuring "a fair, equal, and significant opportunity to obtain a high-quality education ..." (NCLB, 2002; Title 1 Section 1001). According to NCLB, academic standards must be the same standards that the state applies to all schools and

children (Section 1111(b)(1)(B)), states must develop and implement a single statewide accountability system (Section 1111(2)(A)), and each state shall demonstrate adequate yearly progress toward enabling students to meet its academic standards (Section 1111(2)(B)). NCLB requires states to build on existing high-quality assessment and accountability systems that are coordinated with programs under prior legislation such as IDEA, the Head Start Act, and the McKinney-Vento Homeless Assistance Act (NCLB, 2002; Paige, 2002). Up to this point, assessing students with significant disabilities was only an issue for special educators because it applied to a small number of students held to a different set of standards. However, school districts can no longer account for these students separately.

Assessment and Accountability Systems

Alternate assessment has been in place only a short time, and the number of students who participate may be as small as one half to two percent (Thurlow et al., 1996; Warlick & Olsen, 1999). This recency contributes to the paucity of literature regarding the impact of alternate assessment on instructional practice and student outcomes. States such as Kentucky and North Carolina, however, have been identified as leaders in educational reform and may offer examples of successful approaches to inclusive assessment and accountability (Quenemoen, Rigney, & Thurlow, 2002). According to Thurlow and colleagues (1996), what they demonstrate and what other states can glean from them is that "... an alternate assessment system has very little value unless the results from this system are integrated into the general accountability system" (Alternate Assessment Issues section, ¶8).

State Implementation of Alternate Assessment

The development and implementation of states' alternate assessments has been incremental. Early alternate assessment frameworks provide evidence of the rush to meet the implementation deadline. Before July 1, 2000, only two states, Kentucky and Maryland, were including all students in their accountability systems though other states, such as Texas and California, had blueprints for similar efforts. Since that time, changes in practice have occurred because of input from educators and other stakeholders, emerging research, and legislative mandates (deFur, 2002; Haigh, 1999; Olsen, 1999a, 1999b; Quenemoen, personal communication, February 7, 2003; Thompson & Thurlow, 2000, 2001; Virginia Department of Education [VDOE], 2000a, 2000b, 2000d, 2001, 2002). Many of these changes show early promise for positive outcomes though evidence of unintended consequences is beginning to emerge, such as the inclusion of students who are not appropriate for the assessment. As states work toward including all students and developing technically sound means for doing so, both policy and practice will benefit from reports on their progress in reaching assessment and accountability goals.

Research on Alternate Assessment Development and Implementation

The majority of research and reports on alternate assessment implementation and status of state progress comes from the National Center on Educational Outcomes (NCEO), located within the College of Education and Human Development at the University of Minnesota. "This organization was established in 1990 to provide national leadership in designing and building educational assessments and accountability systems that appropriately monitor educational results for all students, including students with disabilities and students with limited English proficiency" (National Center on

Educational Outcomes [NCEO], n.d., Overview section, ¶1).

Since its inception, NCEO has focused specifically on (a) working with states and federal agencies to identify important outcomes of education for students with disabilities; (b) examining the participation of students in national and state assessments, including the use of accommodations and alternate assessments; (c) evaluating national and state practices in reporting assessment information on students with disabilities; (d) bridging general education, special education, and other systems as they work to increase accountability for results of education for all students; and (e) conducting directed research in the area of assessment and accountability (NCEO, n.d.). Their work is credited for being “a major component in the evolution of federal policy related to the participation of students with disabilities in assessments” (Ahearn, 2000, p. 8).

Role of NCEO in Establishing Core Principles

NCEO’s director, Martha Thurlow, and her colleagues have documented the progress of researchers, policymakers, and stakeholders through this era of educational reform. In addition to these research efforts, NCEO has worked closely with personnel at the national, state, district, and school levels to identify challenges that educators and students face in standards-based reform. Collaborating with such organizations as the Council of Chief State School Officers (CCSSO) and the National Association of State Directors of Special Education (NASDSE), NCEO’s work led staff to the opinion that early guidelines for participation of students with disabilities in large-scale assessments, originally developed by Elliott, Thurlow, and Ysseldyke (1996), were due for revision. Subsequently, Thurlow and colleagues have developed what they consider core principles and characteristics of inclusive assessment and accountability systems based on a decade

of their research (Quenemoen, Thompson, Thurlow, & Lehr, 2001; Thurlow, Quenemoen, Thompson, & Lehr, 2001). These principles are derived from their documentation of accountability and assessment systems and on the comments and reviews of numerous stakeholders (Thurlow et al., 2001). Further, these authors believe that in order to enhance the positive consequences of inclusive assessment and accountability systems and minimize the negative ones, educators and decision-makers must examine the underlying assumptions on which their assessments are based and move toward more inclusive assessment systems (Thurlow et al., 2001). According to Quenemoen, Thompson et al. (2001),

The purposes of the “Principles and Characteristics of Inclusive Assessment and Accountability Systems” is to focus and clarify stakeholder discussion on essential components of inclusive systems state by state and district by district and to provide an impetus for revisiting basic assumptions and beliefs about emerging state and district systems. (pp. 3-4)

Though consistent with the legal requirements of the NCLB and its predecessors, Improving America’s Schools Act of 1994 (IASA 94) and IDEA 97, the principles generated by the staff at the NCEO are intended to go beyond them and offer characteristics of best practice (Quenemoen et al., 2001; Thurlow et al., 2001). Broadly, the six core principles address the following themes: (a) all students, (b) decisions, (c) reporting, (d) accountability, (e) improvement, and (f) inclusive school reform.

Statement of the Purpose

The purpose of this study was to examine alternate assessment policy from six states with an emphasis on the existence of the core principles of inclusive assessment

and accountability systems reported in the literature. To accomplish this, the researcher first searched for evidence of the core principles espoused by emerging policy, research, and best practices in state documents and reports. Next, the researcher compared and contrasted state policy and records. Finally, the researcher identified additional constructs beyond those core principles reported in the literature.

Research Questions

State data and documentation exist to provide a means of examining the status of alternate assessment policy and the existence of the core principles and characteristics described in the literature (Quenemoen, Thompson et al., 2001; Thurlow et al., 2001). These principles are evidenced by (a) inclusion of all students in assessments; (b) assessment system decisions based on clearly articulated participation guidelines; (c) scoring and public reporting of all results; (d) inclusion of all results in accountability indices; (e) improvement in assessment and accountability systems through monitoring, evaluation, and training; and (f) written policies that reflect the belief that all students be included in assessments. This study analyzed data collected from various sources to answer the following questions:

1. Are the core principles and characteristics of inclusive assessment and accountability systems reported in the literature present in legislation, policies, and procedures that govern state alternate assessment programs?
2. What similarities exist across state alternate assessment policy?
3. What differences are evident in state alternate assessment policy?
4. What additional themes emerge from analysis of the state legislative, policy, and procedure documents that drive alternate assessment programs?

Significance of the Study

No longer are advocates for students with disabilities the only ones calling for inclusive accountability systems. Because yearly progress reports must include the performance of all students, policymakers, educational leaders, and the community feel the sense of urgency that until now only families and some educators felt regarding the participation of students with disabilities in assessment and accountability systems. Alternate assessment is now part of the general education vocabulary. The assessment, defined by IDEA 97, has only been conducted in the majority of states for four years. Therefore, few states have begun to evaluate their programs. Nonetheless, NCLB calls for the use of alternate assessment as one means of ensuring every student's participation in the assessment program and subsequent system accountability. This requirement poses significant challenges to educational leaders for policy formulation and subsequent implementation.

The majority of the data on alternate assessment is descriptive and addresses states' implementation of the IDEA mandate. States must begin to evaluate their policies against emerging research and best practice to enhance positive consequences and minimize the negative consequences for all students (Thurlow et al., 2001). Such efforts will further enable states to determine if their alternate assessment programs are aligned with emerging best practice. This study offers valuable information to assessment and accountability administrators as well as policy researchers for the following reasons:

1. Aside from literature on state implementation of alternate assessment, there are no published accounts of the congruence of these programs with emerging best practice. This study offers such information, which is important because states

have spent the past year and a half trying to make sure their alternate assessment programs “fit” the mandates of NCLB and must now respond to the utility of the assessment as an indicator of student achievement.

2. States have begun to evaluate their compliance with IDEA mandates for alternate assessment. However, they have only recently begun to address the assessment and accountability requirements of NCLB, many of which are formula-driven or quantitative. This study provides qualitative information to characterize the content of the policies that shape alternate assessment. Such information is important, as states need both empirical and descriptive data on current practice to inform their decision making during this shifting policy environment.
3. By examining the policies and procedures that govern states’ alternate assessment programs, one can glean a sense of how policymakers have operationalized the mandates of NCLB. Although state constitutional provisions for education differ, information obtained in this study illuminates state policies and procedures that provide acceptable responses to constituent and legislative mandates. Such information could be beneficial to state leaders in determining if policies are aligned with best practice cited in the literature.

This study is especially timely subsequent to the issuance of the final regulations implementing Title I of NCLB (NCLB, 2002). As a result, the inclusion of students with severe to profound disabilities in accountability systems is at the center of attention at federal and state levels. The areas of NCLB with particular relevance to the present study include State Accountability Systems and Adequate Yearly Progress (AYP) because they have forced states to focus on how they will include students with severe to profound

disabilities in their accountability indices. Because of the large number of comments received, including many reflecting a misunderstanding of the standards for AYP, the Secretary of Education issued a Notice of Proposed Rule Making (NPRM) asking for additional public comment on the proposed rule regarding the use of alternate achievement standards for students with the most severe disabilities. This area will likely receive even more attention as states begin to fulfill the relevant legislative requirements.

Definitions

Accountability system. Used to evaluate whether, and to what degree, individuals meet standards or expectations. In educational contexts, accountability systems provide a systematic method for assuring internal and external audiences that schools are moving in the right direction (Thurlow, Elliott et al., 1998).

Adequate yearly progress (AYP). State-established definition of annual objectives to measure progress of schools and districts to ensure that all groups of students (including low-income students, students from major racial and ethnic groups, students with disabilities, and students with limited English proficiency) reach proficiency. In order to make adequate yearly progress, schools must test at least 95% of their students in each of the above groups (NCLB, 2002).

Alternate achievement standards. Described by the U.S. Department of Education as an expectation of performance that differs in complexity from a grade-level achievement standard. These standards may be used to measure the achievement of students with the most significant cognitive disabilities although only one percent of the scores based on these standards can be used in the calculation of AYP (34 CFR Part 200).

Alternate assessment. An assessment defined and mandated by IDEA 97 that is used to measure the performance of students with significant disabilities who are unable to participate, even with accommodations, in general large-scale assessments using the same academic achievement standards as all other students.

Assessment. Refers to the process of measuring learning against a set of standards (Thurlow, Elliott et al., 1998).

Content standards. Statement or description of the knowledge and skills in a content area (e.g. mathematics, science, or social studies) that students are expected to learn (Elliott et al., 2001).

Core principles. Essential components of inclusive assessment and accountability systems that are consistent with, yet go beyond, legislative mandates (Thurlow et al., 2001).

Gray area of assessment. Areas of concern that exist in a state or districtwide assessment program when a student does not meet the participation criteria for the general assessment, without or with accommodations, or the alternate assessment (Almond, Quenemoen, Olsen, & Thurlow, 2000).

Indicators. Characteristics that describe the presence of NCEO's core principles for inclusive assessment and accountability systems (Quenemoen, Thompson et al., 2001).

Large-scale assessment. An approach to testing whereby an entire population of students is administered an achievement test as part of an accountability system (Elliot et al., 2001).

Out-of-level testing. Assessing students in one grade level using versions of tests that were developed for students in other grade levels. Typically, the tests are designed for lower grade levels (Heumann & Warlick, 2000).

Performance standards. Also referred to as achievement standards, reflect how well children must demonstrate what they know and are able to do (Thompson & Thurlow, 2003). Expected achievement levels for state- and district-defined content standards, they describe the quality of performance expected for proficiency on the content standard (Thompson et al., 2001).

Test. A structured performance situation that can be analyzed to yield numerical scores from which inferences are made about how individuals differ in the construct measured (Gall, Borg, & Gall, 1996).

Limitations, Delimitations, and Assumptions

Limitations

Rudestam and Newton (2001) described limitations as "... restrictions in the study over which you have no control" (p. 90). The researcher was faced with several such limitations in this study. A major limitation was posed by the newness of the implementation of alternate assessment. Because this measure has only been in place for less than four years in the majority of states, the body of literature in this area is relatively small. Studies on alternate assessment are emerging, but many of them describe its status rather than its impact on instruction and student learning. Additionally, the paucity of data limits judgments regarding the validity of the results from the present study because of the inability to substantiate the researcher's conclusions. A final limitation was the degree to which state documentation exists and was accessible. In addition to the size of

the sample number of documents being small, not all state documents pertaining to alternate assessment and accountability were available to be reviewed and analyzed. The sample, however, was purposefully selected, and the author attempted to support all generalizations.

Delimitations

Delimitations imply limitations or boundaries on the research design that the researcher has imposed deliberately, but which may limit generalizability (Glatthorn, 1998; Rudestam & Newton, 2001). This study examined policy from six states in the mid-eastern United States. This limited examination precludes making assumptions about the alignment of alternate assessment policy with best practice across the remaining states. Further, the close proximity of the states surveyed may yield little variation in data, especially since state administrators may have assisted each other in developing their policies and procedures. Further, content analysis was limited by the inherent difficulties in developing categories and the criteria for assigning content units to categories (Boyatzis, 1998). This study used codes generated from prior research to attempt to reduce the problems associated with category development.

Assumptions

Several assumptions guided this research. First, it was assumed that there was a correspondence between pre-established core principles found in the literature and inclusive assessments and accountability systems. Second, it was assumed that the documents identified for inclusion represented those most reflective of the regulations, policies, and procedures that govern and dictate states' alternate assessment programs.

Berger and Luckman (cited in Neuendorf, 2002) offered that “there is no such thing as true objectivity – ‘knowledge’ and ‘facts’ are what are socially agreed on” (p. 11). In this study, the author’s reality of alternate assessment was greatly influenced by exposure to the process at the Virginia Department of Education during an administrative internship. This was considered in designing a research method with sufficient rigor to increase the quality of the conclusions drawn.

CHAPTER II - Review of the Literature

This chapter includes a review of special education reform, legislation related to assessment and accountability systems, policy implementation and its impact on practice, inclusive accountability systems, implementation of alternate assessment, and core principles of assessment and accountability. Personal contact with authors of seminal studies (e.g., K. Olsen, E. Ahearn, and R. Quenemoen) provided the initial and subsequent direction for the review. Information for the literature review was obtained using not only local library holdings or those obtained via interlibrary loan, but also several databases. The following database descriptors were used: *alternate assessment, academic assessment, disabilities, educational reform, accountability systems, academic standards, reform efforts, access to education, inclusive schools, standards-based reform, law and legislation, special education, educational change, and education outcomes*. A major criterion of the review was to identify articles and documents that discussed the assessment of students with significant disabilities and inclusion of assessment results in state accountability systems. The goal of this review was to locate and present information on the impact of education reform and policy on the inclusion of students with disabilities in assessment and accountability systems.

Impact of Special Education Reform on Assessment Systems

Special education reform is most notably linked to general education reform. In the context of inclusive accountability systems, significant change emerged in the 1980s

and 1990s (Kleinert & Kearns, 2001; Turnbull, Turnbull, Shank, & Leal, 1999).

Widespread public concerns that the educational system was inadequate prompted the U.S. Secretary of Education, Terrel H. Bell, to form the National Commission on Excellence in Education in 1981. This Commission's 1983 report, *A Nation at Risk: The Imperative for Educational Reform* (Nation At Risk, 2003), urgently called for both immediate and sustained improvement. Specifically, a recommendation to adopt more rigorous and measurable standards and higher expectations for academic performance were identified. In so doing, the report called for increased accountability at the local and state level, an increase in the amount of homework assigned, more courses, and higher standards for teachers as well (Kleinert & Kearns, 2001; Smith & O'Day, 1991). These initial changes did little to change the content of instruction, or directly involve stakeholders (Smith & O'Day, 1991). In what they refer to as a "second wave," occurring in the late 1980s to early 1990s, Smith and O'Day (1991) noted a push for greater parent involvement, greater student and teacher engagement, and a more challenging curriculum using a decentralized, bottom-up approach.

The changes of the early 1990s also ushered in national content standards. Students with disabilities were excluded from these standards-based assessments (National Research Council [NRC], 1999; Thurlow et al., 2001). However, a positive outcome was the belief that all students could learn and that high standards could be met by identifying what students have to know and be able to do, including students with special needs (Thompson, Quenemoen, Thurlow, & Ysseldyke, 2001; Thurlow, Elliott et al., 1998). At the same time, however, an increase in the rates of referral to special education as well as an increased retention in grade of non-special education students was

observed (Kleinert & Thurlow, 2001). Such practices enabled states to maintain high test scores that might be otherwise reduced by including these low performing students.

Schools continue in 2004 to operate in a standards-based reform environment that defines what students should know and be able to do (content standards) as well as what level of performance is necessary to demonstrate mastery (performance standards) (Ahearn, 2000; McLaughlin, Henderson, & Rhim, 1998; Ysseldyke, Krentz, Elliott, Thurlow, Erickson, & Moore, 1998). Generally, the idea of standards-based reform suggests that student achievement will rise if states (a) set high standards for student performance; (b) develop assessments that measure student performance against the standards; (c) give schools the flexibility they need to change curriculum, instruction, and school organization to enable their students to meet the standards; and (d) hold schools strictly accountable for meeting performance standards (NRC, 1999, p. 15). As intended, this movement has produced sustained emphasis on assessment and accountability in both general and special education.

Legislative Background

In 1989, shortly after he took office, President George Bush invited the Nation's 50 governors to attend an Education Summit to discuss the current condition of education and what course of action might be adopted to reverse the trend toward mediocrity (U.S. Department of Education, 1994, Part III, ¶5). Although a consensus was reached on the nature of existing educational problems and the strategies necessary to solve these problems, implementation of these strategies failed. Nonetheless, the Educational Summit of 1989 set the stage for education reform to include higher expectations, rigorous standards, and assessment of progress for all students. Subsequently, Goals 2000, the

Elementary and Secondary Education Act (ESEA), the Individuals with Disabilities Education Act Amendments of 1997 (IDEA 97), and most recently the No Child Left Behind Act of 2001 (NCLB) reinforced these expectations (Quenemoen, Lehr et al., 2001; Thurlow et al., 2001).

Among the laws driving current practice is also the Improving American Schools Act of 1994, which reauthorized the ESEA. This reauthorization required that students in Title I-funded school programs meet the same high expectations as other students (Council for Exceptional Children [CEC], 2000; Thompson et al., 2001). Students with disabilities are specifically included in the language.

On July 1, 2000, the manner by which students with severe to profound disabilities were assessed changed. Henceforth, under the IDEA 97, all states were required to have in place an alternate assessment program that included all students previously ignored in state- and districtwide assessments. The amendments of IDEA 97 further clarified Congress' intent that all students, including those with disabilities, be held to challenging standards. According to Thompson et al. (2001), IDEA 97 focused states' and districts' attention on the challenges of full participation of students with disabilities in assessment systems that would improve outcomes for all students. Two of the most explicit and fundamental changes of IDEA 97 were the requirements that IEP teams address (a) how students with disabilities participated and progressed in the general curriculum and (b) how the learning of students with disabilities was measured and reported (Kleinert & Thurlow, 2001). Thus, students with disabilities would now be required to participate in the general assessment (a) with or without accommodations or with modifications, or (b) to participate in an alternate assessment.

Three major reasons emerged for including these students in the accountability system: (a) to provide an accurate picture of student achievement, (b) to make accurate comparisons between school districts, and (c) to ensure that students with disabilities benefit from reforms (CEC, 2000). Specifically, IDEA 97 states that “children with disabilities be included in general State and district-wide assessment programs, with appropriate accommodations, where necessary” (612(a)(17)(A) and

as appropriate, the State or local educational agency: (i) develops guidelines for the participation of children with disabilities in alternate assessments for those children who cannot participate in State and district-wide assessment programs; and (ii) develops and, beginning not later than July 1, 2000, conducts those alternate assessments. (612(a)(17)(A)(I-ii)

To comply with the requirements of Section 300.138 of IDEA 97, each state must demonstrate that individuals with disabilities are included in its accountability system, and must have in place guidelines for the participation in alternate assessments of those students with disabilities who are not able to participate in state- and districtwide assessments (Thompson & Thurlow, 2000). This legislation further requires that the state educational agency make results available to the public and report them with the same frequency and detail as it does those of children who are not disabled. Additionally, these reports must include both aggregated and disaggregated data reflecting the performance of all students in the system and, separately, the performance of students with disabilities (Thompson & Thurlow, 2000). In the past, results for students with disabilities were excluded in the data collection that measured student progress (Browder, Ahlgrim-Delzell et al., 2002; Vanderwood, McGrew, & Ysseldyke, 1998). Ysseldyke and Olsen

(1999) and others (Burgess & Kennedy, 1998; Krentz, Thurlow, & Callender, 2000; NRC, 1999; Thompson et al., 2001; Thurlow & Johnson, 2000) note that when students are out of sight in the assessment and accountability process, they are out of mind when decisions are made regarding reform. By reporting accurate information on students with disabilities, states ensure that they are represented in the accountability system.

Public Law 107-110 was enacted January 8, 2002 “to close the achievement gap with accountability, flexibility, and choice, so that no child is left behind” (20 USC 6301). This Act, the No Child Left Behind Act of 2001 (NCLB, 2002), which reauthorizes the ESEA calls for increased accountability via challenging standards, annual assessment, and evidence of yearly progress. To ensure that states, school districts, and schools are meeting their objectives the law calls for separate measurable objectives for various subgroups, including students with disabilities ((Section 1111(2)(C)(v)(II)(cc)).

According to Title I of NCLB, in order to demonstrate adequate yearly progress not less than 95% of each group of students identified in the Act take assessments with or without accommodations, or participate in an alternate assessment. The Act was enacted to help states build on the accountability systems already in place, not to initiate new systems. Further, it complements assessment and accountability programs established under the guidelines of other legislation such as IDEA. NCLB calls for increased accountability for states, school districts, and schools as a focus area. Specifically, Section 1111(B)(2)(a) of the Act states:

Each State plan shall demonstrate that the State has developed and is implementing a single statewide State accountability system that will be effective

in ensuring that all local education agencies, public elementary schools, and public secondary schools make adequate yearly progress as defined under this paragraph.

The Education Commission of the States (ECS), created to improve public education, is one organization that monitors states' progress toward meeting the requirements of NCLB. In 2003, the organization compiled a database containing information from each state that represented state laws, departmental regulations, board rules, directives, and practices related to 40 requirements across seven major sections of the NCLB legislation (Education Commission of the States [ECS], n.d.). Specific to this research were the sections on Standards and Assessment, and Accountability (Adequate Yearly Progress [AYP]). The ECS database reported states' progress toward meeting the requirements of NCLB using the following descriptors: (a) appears to be on track, (b) appears to be partially on track, (c) does not appear to be on track, and (d) unclear or data not available.

According to ECS, by 2003 the six states that were analyzed in the present study appeared to be on track, showing evidence of policies that ensured the inclusion of 100% of the students with disabilities in their state assessments. Accountability was the second section reviewed, along with three of its requirements: (a) a single accountability system, (b) accountability for all subgroups, and (c) 95% of students in all subgroups assessed. The requirement of a single accountability system posed a challenge for state educational agencies. ECS found that Maryland, West Virginia, Kentucky, Tennessee, and North Carolina appeared to be on track in meeting this requirement while Virginia appeared to be partially on track. ECS described accountability for all subgroups as all public schools

and LEAs being held accountable for the achievement of individual subgroups. All of the states appeared to be on track, except for Kentucky, which was partially on track. For the last requirement reviewed, 95% of students in all subgroups assessed, Virginia, North Carolina, and West Virginia were reported to be on track whereas Maryland, Kentucky, and Tennessee were noted to be partially on track.

In the midst of this monitoring, the Department of Education has issued a final regulation amending a prior regulation to now permit the use of alternate achievement standards for students with the most severe cognitive disabilities. This regulation, effective January 8, 2004, allows states to “give equal weight to proficient and advanced performance based on the alternate standards in calculating school, district, and State AYP, provided that the number of proficient and advanced scores does not exceed 1.0 percent of all students in the grades tested at the State or LEA level” (34 CFR Part 200, p. 68699).

Some states were already using different standards or assessments with this population that were not at grade level. They now must revisit their use to ensure that these assessments are aligned with state content standards and yield results in both reading/language arts and mathematics. Further, states that were trying to fit these students into their existing assessment program have the opportunity to make modifications based on this new policy.

The Secretary of Education has indicated that he welcomes comments regarding how this policy is working over time to determine if revisions are warranted. As may be the case with policy implementation, once states act on their interpretation of a regulation, it may result in a new or modified view of the policy.

Policy Implementation

Policy is developed close to the top of the political system but concerns regarding implementation occur at the state and local levels. Researchers have described public policy in numerous ways, using descriptors such as dynamic and value-laden, or referring to it as the output of a political system or the process political systems use to address public issues (Fowler, 2000; Rist, 2000; Spring, 1998). All policy includes expressed intentions, although statutes are often worded in general terms or contain vague, ambiguous language (McDonnell, McLaughlin, & Morison, 1997; Yanow, 2000). The details of these laws are provided by the governmental agencies such as state departments of education or state boards of education, and at times regulations must be amended to provide clarification.

Spring (1998) makes a distinction between the politics of policy and the politics of implementation. According to him, national education policy is the result of the struggles between politicians and various interest groups. Implementation, on the other hand, involves give-and-take among, politicians, interest groups, and bureaucrats. Yanow (2000) contended, "Policy implementation is the social construction of a reality: it is a process of meaning-making through interpretation" (p. 222). Subsequently, it is important to examine whether those who implement state policy have the same understanding of the intent of the policy as its creators (Fowler, 2000; Rist, 2000; Yanow, 2000). Fowler suggested that policies are always changed in the process of implementation because implementers make their own meaning. Yanow goes further to suggest that "once an implementer interprets a policy and acts on that interpretation a 'reader' of that interpretation is no longer dealing with the original policy" (p. 230). Leaders must not

only determine if state alternate assessment policy as described in IDEA 97 has new meaning within the framework of NCLB but also if their implementation of these policies is consistent with the intent of the original mandates and has not been significantly altered during interpretation. To do so policymakers must evaluate assessment programs to determine if they are yielding the outcomes for which the policy was designed.

NCLB suggests several critical elements to ensure that schools are held accountable for outcomes in order to provide the best education for all students. These include (a) academic content standards, (b) academic achievement standards, and (c) assessments aligned with these standards. With these in place, the result should be inclusive assessment and accountability systems that yield positive outcomes for all students.

Inclusive Accountability Systems

“The most frequently cited defining element of the current educational reform movement is accountability” (Ahearn, 2000, p.1). Central to current efforts are both assessment and accountability reforms that include an increase in the number and types of assessments states use and accountability reforms that focus on school districts, accreditation, test scores, and rewards and sanctions tied to student performance (McLaughlin et al., 1998). Erickson, Ysseldyke, Thurlow, and Elliott (1998) contended that educational accountability includes three factors: (a) expectations for learner outcomes or content standards, (b) some form of comparing these expectations with actual outcomes, and (c) positive and negative consequences of student outcomes for school systems (p. 7).

Although the number of students participating in the alternate assessment

represents a small percentage of a state's overall student population, researchers have suggested that results will influence curricular decisions and improve instructional practices (Browder, Flowers et al., 2002; Ford, Darvern, & Schnorr, 2001; Kleinert, Kennedy, & Kearns, 1999). Further, these data are part of the evaluation of school performance and may be figured into a school's accountability index. Frequently cited guiding principles for alternate assessment include that schools are accountable and must have high expectations for all students and that assessments must yield reliable and valid information that leads directly to student learning and improved instruction (Ahearn, 2000; Bolt, Krentz, & Thurlow, 2002; Linn, 2001; Quenemoen et al., 2002; Ysseldyke, Thurlow, Kozleski, & Reschly; 1998).

If accountability systems are to be meaningful, assessments must align with standards and curricula (Browder, Flowers et al., 2002; La Marca, 2001; Ysseldyke et al., 1998). As such, standards and curricula must be the result of decisions about what students have to know and be able to do. In the past, such systems were separate; one set of results was reported for students taking the general state- or district-wide assessment, another was reported for students with significant disabilities.

Implementation of Alternate Assessment

Before 2000, very little research had been conducted on alternate assessment. However, there was a research base for alternate assessment. Thus, various descriptions of strategies for testing students with disabilities existed in the literature (Elliott et al., 1996; McDonnell et al., 1997; Thurlow, 1994; Thurlow, Elliott et al., 1998). Alper, Ryndak, and Schloss (2001) note, however, that from its beginnings the assessment of students with disabilities has been faced with criticism and controversy. Nonetheless, the

federal requirement mandating that school districts implement alternate assessment by July 1, 2000 is less at issue than the larger question of how to effectively and efficiently implement this requirement (Kleinert, 2001; Kleinert & Kearns, 2001).

Elliott et al. (2001) delineate eight issues that influence the implementation and meaningful use of alternate assessment. The first of these is alignment with content standards. In their review of alternate assessment practices in 11 states, Elliott and colleagues found that alternate assessments were well aligned with content standards (also known as academic or learning standards). As alternate assessment requirements took effect, Thompson and Thurlow (2000) found that these measures assessed (a) general education standards, (b) a subset of general education standards, (c) standards in addition to general education standards, or (d) different standards from general education. Although IDEA does not specifically state that alternate assessments must be based on the general curriculum, many contend that it is not sound practice to require that students with disabilities be included in the general curriculum and then be assessed using standards outside of this context (Browder, Flowers et al., 2001; Elliott et al., 2001; Kleinert & Thurlow, 2001; Thompson et al., 2001; Warlick, 2001).

Scoring of evidence is identified as a second issue that affects the meaningful use of alternate assessment. The nature of the score depends on the nature of the assessment (Roeber, 2002). With variation in alternate assessments across states, rating scales and rubrics also vary across states. Elliott et al. (2001) reported four similarities. Among these are the counts of how frequently a student demonstrates a task or performs a skill, the amount of support a student needs to complete a task, the generalizability of the skill, and the accuracy and quality with which the student can perform a task or demonstrate a skill.

Elliott and colleagues (2001) suggest that the third issue is time and timing. This includes the need to collect information about student learning over a period of time to make reliable judgments. Further, educators are generally bound by reporting deadlines. These two aspects of the alternate assessment process may pose initial challenges until teachers become accustomed to shifting data collection requirements and timelines.

The fourth issue is parental involvement. Not only must parents receive assessment results, they must also be part of the assessment process. Thompson and Thurlow (2001) reported that 44 states involved parents in the development of their alternate assessment. State documents further support parental involvement. Indiana's alternate assessment guidelines, for example, indicate that in all cases the parent is a member of the team that conducts the alternate assessment.

The fifth issue is the reliability and validity of results. Questions regarding reliability and validity have existed since the development phase of alternate assessment and continue as the majority of states use some form of portfolio, checklist, or student work sample (Quenemoen, Thompson, Thurlow, & Olsen, 1999; Thompson & Thurlow, 2001; Ysseldyke, Olsen, & Thurlow, 1997).

Out-of-level testing is the sixth issue. This practice is not prohibited under IDEA, according to Heumann and Warlick (2000), and recent studies show evidence that its use has increased from 1997 to 2001 (Thompson & Thurlow, 2001). Specific concerns include (a) out-of-level testing may assess different content standards than would be assessed at a specific grade level and (b) scores from the results are not comparable with those from other assessments and therefore cannot be aggregated (Elliott et al., 2001; Quenemoen et al., 1999).

A seventh issue involves the storage of all of the information collected with alternate assessment. Depending on the length of time documents must be kept and the format of the assessment, the attention this issue receives will vary. Indiana, for example uses electronic portfolios therefore minimizing this concern, whereas states such as Idaho maintain student files for five years (Elliott et al., 2001).

The final issue Elliott and colleagues address is the reporting results of alternate assessments. This area is receiving increased attention now that states will be held accountable for assessing all students and including all results (Bechard, 2001; Quenemoen et al., 2002; Wiener, 2002). Though Elliott and colleagues (2001) consider scoring and reporting to be related, they contend that reporting represents the “bottom line” (p. 93). Because the focus of alternate assessment varies and the information collected may be different for each student, aggregating the results and making comparisons across groups raises questions and warrants attention (Bechard, 2001; Hill, 2001; Linn, 2001; Quenemoen et al., 2002; Quenemoen & Thurlow, 2002).

The purpose of alternate assessment is to improve instruction and results for students with disabilities (Kleinert & Kearns, 2001; Thompson et al., 2001; Ysseldyke & Olsen, 1999). Yet, states’ practices vary across all of the dimensions identified by Elliott et al. (2001) and others (Browder, Ahlgrim-Delzell et al., 2002; Burdette & Olsen, 2000; Thompson & Thurlow, 2000). One of the challenges facing states in their implementation of alternate assessment is that IDEA 97 does not specify guidelines for participation, approaches to be used, measures of proficiency, scoring, or reporting results of the assessment. As a result, a great deal of variability exists across states in their alternate assessment practices (Browder, Ahlgrim-Delzell et al., 2002; Hager & Slocum, 2002;

Thompson & Thurlow, 2000). These are all issues that states have considered in developing and implementing their alternate assessment programs. A brief overview of the programs in KY, MD, NC, TN, VA, and WV will provide a glimpse of how they have dealt with these issues.

Kentucky

The Kentucky Alternate Portfolio (KAP) assessment was the result of comprehensive Kentucky Education Reform Act (KERA) in 1990, making it the first alternate assessment used in this country for students with disabilities (Kearns, Kleinert, & Kennedy, 1999; Ysseldyke et al., 1996). With the implementation of this sweeping legislation came school control of curriculum and instructional decisions as well as accountability for student achievement. Not only does the assessment measure what students know and are able to do, it also measures the degree to which schools and programs implement the research-based strategies its developers have identified as important for students with disabilities.

The assessment was the product of a group of teachers, parents, administrators, university personnel, and education consultants who met with the Kentucky Department of Education and assessment contractors to develop the process that would be used. According to 2002-2003 Kentucky Alternate Portfolio Project information (KAP Project, n.d.), the number of students participating in this alternate assessment has remained fairly constant at approximately .6% of the total school population, or between 900-1,000 students per accountability year.

In 2002-2003, students who might be appropriate for the KAP were evaluated based on seven criteria during a yearly Admissions and Release Committee meeting.

Students who did not meet all of the criteria could not participate in the alternate assessment program; instead, they took part in the Commonwealth Accountability Testing System (CATS), with or without modifications/accommodations. The seven criteria for participation included:

1. The student has a current Individual Education Plan
2. The student's demonstrated cognitive ability and adaptive behavior itself prevents completion of the regular course of study even with program modifications and/or accommodations
3. The student's current adaptive behavior requires extensive direct instruction in multiple settings to accomplish the application and transfer of skills necessary in school, work, home, and community environments
4. The student's inability to complete the course of study may NOT be the result of excessive or extended absences; or it may NOT be primarily the result of visual or auditory disabilities, specific learning disabilities, emotional-behavioral disabilities, or social/cultural/economic differences
5. The student is unable to apply or use academic skills at a minimal competency level in natural settings (i.e. home, community, or work site) when instructed solely or primarily through school-based instruction
6. The student is unable to acquire, maintain, or generalize skills, and demonstrate performance without intensive, frequent, and individualized community-based instruction
7. The student is unable to complete a regular diploma program even with extended schooling and program modifications/accommodations.

(KAP Project, n.d.)

The KAP, which predates IDEA 97, is based on a set of six learning goals, 54 academic expectations, and a program of studies, and all students in Kentucky must be making progress toward these goals and expectations. After researching several formats, the portfolio was identified for use with students with moderate and significant disabilities who are unable to participate in the CATS. The KAP is primarily a collection

of student work compiled over a period. Additionally, it may include peer and teacher notes, data and graphs, video or audiotapes, and photographs. Entries are based on five content areas that vary depending on grade level, and from which a holistic score is derived. They include (a) language arts, (b) math, (c) science, (d) social studies/vocational, and (e) arts and humanities or health and physical education (KY Department of Education, 2002).

In addition to these content areas, the 2002-2003 assessment also included the following components: (a) table of contents, (b) letter to the reviewer reflecting the portfolio process and content, (c) letter from the parent/guardian validating the contents, (d) an individualized daily student schedule with a description and evidence of its use, (e) a formal resume (grade 12), and (f) evidence of the student's mode of communication.

In the 2002-2003 academic year, individual teachers scored the alternate portfolio during regional scoring sessions. Each was scored three times; the first two scores were compared, and the third was used to resolve areas of disagreement. All teachers who assisted with compilation of the alternate portfolio were required to participate in the scoring process. However, teachers did not score their own students' work (KY Department of Education, 2002). Scoring was accomplished using a rubric that addressed six dimensions: Standards, Performance, Settings, Support, Social Relationships, and Self-Determination. Alternate assessment scores were used to evaluate schools and programs and were factored into the school accountability index across the content areas covered by the general assessment.

Maryland

Implementation of Maryland's alternate assessment, the Alt-MSA (formerly the

Independence Mastery Assessment Program [IMAP]), followed closely behind that of Kentucky. Much like Kentucky, the state was pressured into developing and implementing a system that would hold programs and students accountable for educational outcomes. In 1989, Maryland's State Board of Education responded to a report by the Governor's Commission on School Performance that noted a lack of information on student and school performance and subsequent accountability (Ysseldyke et al., 1996, 1997). What followed was a state school performance program that included an assessment of all students, making Maryland, along with Kentucky, front-runners in inclusive assessment and accountability systems.

Through an IEP process, students for whom the Maryland School Assessment (MSA) is deemed inappropriate and who are pursuing an alternate course of study participate in the Alt-MSA (MD State Department of Education [MSDE], n.d.). Participation in the Alt-MSA is based on the outcomes that the student is pursuing, is content standards or life-skills oriented, and includes three components: (a) performance tasks, (b) parent survey, and (c) student portfolio (Ysseldyke et al., 1996). The program uses a combination of standards and extensions of the standards that complement the general curriculum and content areas but focus on life skills. The Alt-MSA is similar to the general assessment in the time of the assessment, frequency, and reporting. Students are assessed at grades 3, 5, and 8, and grade 10 for math and reading. Alt-MSA scores were reported in 2002 as an aggregated score; however, with recent revisions to the 2002-2003 assessment, separate scores for math and reading will enable students participating in the Alt-MSA to be reported as basic, proficient, and advanced. Scores going forward will be combined with those from the MSA to determine AYP (MSDE, 2003).

North Carolina

Like the majority of states, the North Carolina alternate assessments were developed in response to the mandates of IDEA 97 and implemented July 1, 2000. The state offers two alternate assessments for students with disabilities: the North Carolina Alternate Assessment Portfolio (NCAAP) for students with severe cognitive disabilities and the North Carolina Alternate Assessment Academic Inventory (NCAAAI) for students who do not benefit from accommodations and are not candidates for the NCAAP. The focus of this study was the NCAAP.

The NCAAP was piloted during the 1999-2000 academic year with students at grades 3-8 and volunteer districts at grade 10 to evaluate its feasibility, validity, and reliability for use of the portfolio to assess students with serious cognitive disabilities (NC Department of Public Instruction [NCDPI], 2000, 2001). Prior to the implementation of NCLB, the state included tasks related to literacy, numeracy, and technology in a student's alternate assessment portfolio. To align with the changes mandated by NCLB legislation, North Carolina now assess students in math and reading in grades 3-8 as well as in writing at grades 4, 7, and 10 (NCDPI, 2003). While these scores are used to determine adequate yearly progress, the state continues to use the NCAAP results to evaluate overall performance of its statewide ABCs Accountability Program (NCDPI, 2001).

The NCAAP specifically addresses four domains that represent an extension of the state's *Standard Course of Study*. They include Communication, Personal and Home Management, Career and Vocational, and Community. The IEP team determines a student's eligibility to participate in the NCAAP based on the following five criteria: (a)

the student has a disability and a current IEP, (b) the student has a serious cognitive deficit, (c) the student is assigned to the third through eighth grade or to grade 10 based on the information in the school information management system, (d) the student's program of study focuses on extensions of the NC *Standard Course of Study*, and (e) the IEP team has determined that the student cannot participate in the statewide assessment even with accommodations (NC State Board of Education, 2003).

The assessment is completed over the academic year and culminates in a collection of students' work (portfolio) based on their IEP goals. Two North Carolina teachers with experience in working with students with significant cognitive deficits score each portfolio. In some instances, professional scorers are also used (NC State Board of Education, 2003). Initially, a Task Rubric is used to determine student performance. Each domain is scored 0-4 across the content areas and is then summed to determine a Total Portfolio Score. Additionally, each student's performance is reported as an Achievement Level and growth indication. The Portfolio Achievement Level is considered the overall functional level for a student and is represented as Level I, II, III, and IV. This level is reported and used in the state's Performance Composite of the ABCs Accountability Program. Students receiving a score of 17-32 (Levels III and IV) are considered to have met the performance standard. The growth indication for a student is reported as showing growth (++), showing no growth (--), or showing minimal growth (+-) in one or more domain (NCDPI, 2002; NC State Board of Education, 2003).

In addition to the student's score, the portfolio receives a Quality Score, based on the Portfolio Quality Rubric of 1-4 from two independent readers. This score addresses the thoroughness of the evidence, linkage of the evidence to the student's IEP,

appropriateness of the tasks evidenced, and clarity of the evidence. After scoring has been completed, each student receives an Individual Student Report that includes a reading score and a math score, as well as a writing score for students in grades 4, 7, and 10. Summary reports are compiled for each school and school system. Reading and math scores are included in the determination of adequate yearly progress (NC State Board of Education, 2003).

Tennessee

A task force was developed in 1998 to begin work on the state's alternate assessment, the Tennessee Comprehensive Assessment Program - Alternate Portfolio (TCAP-Alt PA). However, like the majority of states, implementation did not occur until July 2000. The assessment is available for students in kindergarten through grade 11, although not required for K-2 if the school district does not assess students in general education at these levels. Students in grades 3-11 who are not able to fully participate in the Tennessee Comprehensive Assessment Program (TCAP), which is based on state content standards, participate in statewide testing and accountability via the alternate assessment, TCAP-ALT.

This assessment is reported to be linked to the general curriculum via a subset of standards in English, math, science, and social studies, which the state refers to as the Tennessee Curriculum Framework: Extended Standards (TN Department of Education [TN DOE], n.d.). The assessment further incorporates functional skills referred to as "Essential L.I.F.E. Learnings." Like the TCAP, content areas are assessed across the dimension of Context, Choice, Supports, Settings, and Peer Interactions and scored from 1-5 for each dimension. Student work is collected during two periods: from the beginning

of the school year to the end of November, and from the beginning of December to the end of the assessment period the beginning of March.

Each portfolio is scored three times by three separate people or pairs of scorers who have been trained on the scoring procedures developed by the Tennessee Department of Education. In addition to the portfolio contents, each submission must contain (a) Table of Contents, (b) a Portfolio Validation Sheet, (c) evidence of the student's mode of communication, (d) the student's schedule, and (e) the required entries based on grade level. Failure to include two or more of these components results in a score of "0" for the portfolio. The score for each content area is determined by adding all of the dimension scores and recording them on the Student Score Summary Sheet. The three Score Summary Sheets are collected and the totals double checked. Once scored, the portfolio is returned to the teacher and the score submitted to the Department of Education. The alternate assessment is scored during the same window of time as that of the general assessment.

Each June, a statewide scoring institute comprised of 100 teachers reviews 10% of the assessments to validate the scores. The student may receive a score of "Below Proficient," "Proficient," and "Advanced." These scores parallel those of the TCAP scores of "Basic," "Proficient," and "Advanced."

Virginia

In accordance with IDEA 97, the 2000-2001 academic year marked the first year that students with disabilities in Virginia participated in the statewide assessment program via an alternate assessment. As noted in the March 31, 2000, Virginia Alternate Assessment Program (VAAP) Field Test Manual (Virginia Department of Education

[VDOE], 2000a) the purpose of the VAAP is to “capture and evaluate the performance of students who have traditionally been exempted from statewide testing programs” (p. 4).

Over the course of two years the Department of Education worked with a statewide Alternate Assessment Steering Committee, made up of members of the Department of Education, Virginia Institute for Developmental Disabilities (VIDD), parents, and school personnel who would develop a set of performance indicators and delivery practices to the Virginia Board of Education. In addition to researching what other states were doing, this group sought input from the Mid-South Regional Resource Center (MSRRC) as well as Inclusive Large Scale Standards and Assessments (ILSSA), which provided a national perspective on alternate assessment. Some of the states initially reviewed and/or consulted included, but were not limited to, North Carolina, Tennessee, West Virginia, Kentucky, and Indiana.

Student performance on the alternate assessment is based on measurable IEP goals linked to the Standards of Learning (SOL) in a variety of settings. This is consistent with the literature on alternate assessment, which suggests that approaches to developing alternate assessment across the states fall into one of five methods, ranging from standards that are based on the general curriculum to those that are developed separately for the alternate assessment with no link to the general standards (Thompson & Thurlow, 2001). The IEP team makes decisions regarding participation in alternate assessment. Participation is determined on an individual basis by the team using the participation guidelines devised during the development of the assessment. As such, students are required to (a) have a current IEP, (b) demonstrate significant cognitive impairments and adaptive skills deficits, (c) demonstrate a present level of performance requiring

extensive direct instruction, (d) require instruction in a variety of settings, and (e) be unable to complete the standard or advanced diploma program (VDOE, 2000c).

Initially, the alternate assessment in Virginia was conducted four times: (a) Elementary I - no later than the school year in which the student is 8 years old on September 30; (b) Elementary II - no later than the school year in which the student is 10 years old on September 30; (c) Middle School - no later than the school year in which the student is 13 years old on September 30; and (d) High School - one year prior to the student's exit year. These guidelines have been revised and the assessment is now conducted for grades 3, 5, 8, and 11 (VDOE, 2002).

Virginia is among the states that use a portfolio-like approach. In Virginia, the actual assessment consists of a Collection of Evidence (COE) that reflects the student's performance on measurable IEP goals. This COE is not a pencil-and-paper test, but consists of products that evidence student performance. For each area of review or entry, pieces of evidence such as pictures, worksheets, journal entries, or videos, reflecting performance are entered into the collection. An entry is required for the four areas that will be scored. These core content areas, which mirror those of the Virginia SOL, include English, math, science (technology), and history/social sciences. Each of these entries is scored across five dimensions using four to six pieces of evidence, the result being a rating of proficient, advanced, or needs improvement. A student may receive a score point of 1, 2, or 3 (3 being the highest) for each of the five dimensions, which reflects the degree to which demonstration of the dimension is evidenced in the COE. For example, to receive a score of 3 on the Student Performance dimension, the COE must demonstrate considerable evidence of student performance of task(s) related to targeted IEP goal(s).

West Virginia

For over 30 years, the West Virginia statewide assessment program has consisted of multiple assessments. However, it was not until the 2000-2001 school year that the alternate assessment was in place for students grades K to age 21, who could not participate in the norm-referenced Stanford Achievement Test (SAT-9) and other assessments used by the state (WV Department of Education, n.d.). Until 2003, the West Virginia Alternate Assessment consisted of data that reflected student performance on specific instructional goals and objectives (IGOs) related to the general education curriculum. With the 2003-2004 administration, the assessment was based on mandated content standards. These include (a) reading; (b) listening, speaking, and viewing; (c) numbers and operations; (d) measurement; (e) science as inquiry; (f) science subject matter/concepts; (g) citizenship; and (h) economics. These data will be collected over three collection periods throughout the academic year and compiled in a Datafolio. The Datafolios will then be sent to the County Test Coordinator who forwards them to the Department of Education where they will be logged in and scored.

To be considered for the West Virginia Alternate Assessment, 2002-2003 participation criteria indicated that a student must meet the following criteria: (a) have an IEP; (b) undergo a multidisciplinary evaluation; (c) have educational performance data that show significant cognitive impairment; and (d) cannot participate in the general statewide assessment, even with accommodations. For students age 14 and older, the IEP team must also determine that the student is unable to meet the graduation requirements for receiving a regular diploma (WV Department of Education, 2003).

Prior to 2003, the Datafolio consisted of an unlimited variety of primary and secondary evidence such as charts and graphs, worksheets, audiotapes, captioned photos, and so on. However, beginning with the 2003-2004 administration, only primary evidence will be used to score the collection, and this evidence must be limited to no more than three pieces per content standard. Although supporting evidence may be submitted, it will not be scored. West Virginia also uses a rubric in the alternate assessment process, but, unlike other states, the rubric will be used going forward to gauge student achievement, the assessments' connection to content standards, and generalized performance.

Like other states, West Virginia revised its alternate assessment for the 2003-2004 administration to keep pace with changing legislative mandates. State administrators have identified several areas that have received attention, including student participation, use of data, Datafolio components, content standards, and the rubric.

Core Principles and Characteristics of Assessment and Accountability Systems

Thurlow and colleagues at NCEO have developed what they consider core principles and characteristics of inclusive assessment and accountability systems based on a decade of research (Thurlow et al., 2001). They report that these principles are derived from their documentation of accountability and assessment systems and on comments and reviews of numerous stakeholders. In a personal communication with Rachel Quenemoen (March 12, 2003), she summarized to the author the process used to design the principles and characteristics as follows:

In 1999-2000, we convened a working group on assessment that we called NCEO affiliates. It included (a) representatives from each of the Federal and Regional

Resource Center networks, commonly known as the Regional Resource and Federal Centers (RRFC); (b) Federal Comprehensive Center representatives, from those who had worked in assessment for students with disabilities; (c) State Department of Education assessment and special education representatives who had worked on the development of their state alternate assessment and accommodations policies; (d) other researchers in the content area; and (e) U.S. Department of Education staff working on inclusive assessment. It also included our official partners from the National Association of State Directors of Special Education (NASDSE) and the Council of Chief State School Officers (CCSSO). We began with a series of working telephone meetings to identify the “big issues” of inclusive assessment and accountability. After we had identified as a group the key issues, we went to Step Two.

During this step, based on initial discussions, NCEO staff developed a series of draft principles and indicators, and submitted them to the listserv discussion group for response. We went through three successive iterations before we determined that we had captured the key issues and format insights from the group. In Step Three NCEO staff refined the draft materials into the principles and characteristics format and the partner checklists, and asked that the members of the NCEO affiliates from the Regional Resource Centers pilot the tools with at least one state in their region. Three of the Regional Resource Centers did so, each focusing on one component of the tool (participation, accommodations, or alternate assessment). We made revisions based on their pilot.

Fourth, one state had requested that NCEO work with their summer leadership seminar to analyze their current system. We asked if they would be willing to, in turn, pilot our principles, characteristics, and checklists, and they agreed. One NCEO staff member worked with their state staff to co-facilitate use of the tools on site, and although they found the actual results of their work helpful, we again refined the tools based on our experience. That led to the fifth step in which the Federal Office of Special Education Programs reviewed our draft to approve it for publication, and following that we published it.

It is against this backdrop that Thurlow and colleagues (2001) contend, that in order to enhance the positive consequences of inclusive assessment and accountability systems and minimize the negative ones, educators and decision-makers must examine the underlying assumptions on which their assessments, including alternate assessment, are based and move toward more inclusive assessment systems.

Though consistent with the legal requirements of the IASA 1994, IDEA 97, and NCLB Act of 2001, these core principles are intended to go beyond them (Thurlow et al., 2001). Broadly, the six core principles address the following themes: (a) all students, (b) decisions, (c) reporting, (d) accountability, (e) improvement, and (f) inclusive school reform. Thurlow and colleagues also describe what they consider characteristics of each principle. A description of each principle and its characteristics follows.

Principle 1: All Students

This principle states: “All students with disabilities are included in the assessment system” (Thurlow et al., 2001, Overview section.). This principle reflects that all students participate in a state’s assessment program in some way - whether via the general

assessment with or without accommodation, or an alternate assessment. This principle is characterized first by evidence that all students who receive educational services in all settings are included. According to Thurlow and colleagues, this includes students in traditional public schools and those who change schools or placements, as well as students receiving federally funded services in nontraditional settings such as home schools, state-operated programs, and the juvenile justice system. A second characteristic is that there are various ways for the assessment system to participate in the assessment program aside from all students taking the same assessment. This should only be allowed, however, to the extent that alternative ways are permitted for other students. For example, testing students with disabilities out-of-level should only be used if it is an option for all students in the assessment program. The third characteristic of this core principle is that exemptions or exclusions from assessment cannot be based solely on the fact that the student has a disability. Again, the policies that apply to exemptions and exclusions should apply to all students in the assessment system and states should maintain data regarding such requests to avoid exemptions of a disproportionate number of students with disabilities.

The inclusion of all students in the assessment process has created ambiguity regarding a variety of contexts. This is primarily due to the differences across states in the approach, participation guidelines, performance dimensions, scoring, and reporting.

Various approaches have been identified by states to collect data for their alternate assessment. These approaches tend to fall into one of the following categories: (a) observation, (b) analysis of existing data, (c) interview or survey, (d) portfolio, and (e) testing/adaptive behavior scale (Hagar & Slocum, 2002; Roeber, 2002; Thompson,

Erickson, Thurlow, Ysseldyke, & Callender, 1999; Thompson & Thurlow, 1999, 2000; Thompson et al., 2001). Over half of the states use a collection of evidence or portfolio assessment. Some states, such as Connecticut and Oregon, use multiple approaches. States are in the process of revisiting the technical soundness of their assessment strategies (Quenemoen et al., 2002).

Principle 2: Decisions

This principle states, “Decisions about how students with disabilities participate in the assessment system are the result of clearly articulated participation, accommodation, and alternate assessment decision-making processes” (Thurlow et al., 2001, Overview section.). The authors describe five characteristics of this principle, of which four apply to alternate assessment. The authors believe this core principle is evident when:

1. Decisions about how students participate are based on the students’ ability to demonstrate what they know and are able to do.
2. IEP teams make participation decisions based on individual students.
3. The IEP team documents assessment participation decisions and rationale for each student, and reviews these decisions a minimum of once a year.
4. Clear and efficient procedures exist for collecting, compiling, and transferring assessment-decision information for each student to state administrators.

Since implementation in 2000, states have struggled with how broadly to define who participates in alternate assessment (Browder, Ahlgrim-Delzell et al., 2002; Warlick & Olsen, 1999). IDEA stipulates that IEP teams must determine how students will participate in the assessment system, whether they will take the general assessment, and

if not, how they will be assessed (Cohen & Heumann, 2001). The phrases “gray area students” and “gray area of assessment systems” emerged as states faced the challenge of including students who did not seem to meet the participation criteria for either the general or the alternate assessment system (Almond et al., 2000; Quenemoen et al., 1999). Quenemoen and colleagues (1999) suggested that it is not the students but the assessments that yield gray areas. States must ensure that no gaps exist in participation guidelines and decisions for the general assessment or alternate assessment, and that students are deemed eligible for one or the other according to stated criteria.

Alternate assessment is intended for students who cannot participate in the general accountability system even with accommodations or modifications. Guidelines for participation vary across states. However, most states agree that alternate assessment involves participation over a period of months. States have begun to revisit guidelines for participation to eliminate existing gaps in the criteria.

Principle 3: Reporting

Thurlow and colleagues (2001) describe this principle as follows: “All students with disabilities are included when student scores are publicly reported, in the same frequency and format as all other students whether they participate with or without accommodations, or in an alternate assessment” (Overview section). The four major characteristics of this principle are as follows. First, all students who receive educational services regardless of placement are accounted for in the reporting system. Second, the number and percentage of students not accounted for in the assessment program are reported and an explanation is provided for non-participants. Third, all scores that are not reported because of technical issues are reported. Last, stakeholders, including parents,

students, teachers, policymakers, and the media, receive reports that provide a clear explanation of the results.

States face numerous challenges in addressing the issue of reporting. The most common method of reporting scores on alternate assessment involves the use of performance dimensions and proficiency levels. A review of states' practices reveals that not only do variations in levels of proficiency exist but also differences in what the alternate assessment actually measures (Roeber, 2002). Some assessments evaluate skill or level of performance whereas others are interested in degree of progress toward some indicator (Kleinert, Haigh, Kearns, & Kennedy, 2000; Thompson & Thurlow, 2000). Yet others measure level of independence while some are interested in the amount of support a student needs. There does seem to be agreement, however, that student measures should involve a variety of contexts and settings.

Student performance is measured in all states, but some states are also interested in system performance (Roeber, 2002; Thompson & Thurlow, 2001). For example, approximately 20 states measure levels of staff support, variety of instructional settings, and whether the assessment is age-appropriate and challenging for students. Further, 12 states measure participation in general education settings, and 9 states measure parent satisfaction (Thompson & Thurlow, 2001).

In some instances the approach is holistic, involving a student's ability to demonstrate a skill across multiple content areas while in others, an analytic approach is used to evaluate each dimension of student performance and/or program opportunities (Roeber, 2002; Thompson & Thurlow, 2001). In states such as Kentucky, Maine, Alaska, and Tennessee, the proficiency levels of the alternate assessment are the same as those

for the general assessment. Other states, such as Arkansas, California, Florida, and Illinois, use different levels of proficiency for alternate and general statewide assessments (Thompson & Thurlow, 2000).

States have used various processes to develop their measures of proficiency and use several systems for scoring and reporting alternate assessment data (Thompson & Thurlow, 2000; Wiener, 2002). Bechard (2001) reported that resulting measures follow one of six models: (a) the same level is used for both the alternate assessment and the general assessment, (b) a different measure is used for alternate and general assessment but both are treated as the same, (c) a different measure is used for alternate assessment and general assessment, (d) proficiency levels overlap for general and alternate assessment, (e) the lowest possible proficiency level is used for alternate assessment, and (f) proficiency levels are not used. Hill (2001) contended that trying to combine alternate assessment results with those from the general assessment is a policy, not a technical, issue.

Although states use a variety of approaches to conduct alternate assessment, Browder, Ahlgrim-DeLzell et al. (2002) found that 90% of the states they evaluated ($n = 42$) used some type of rubric to score student performance. The remaining scoring methods (10%) included calculating a total score based on percentage of IEP goals met, scoring a pen-and-pencil test, and using a teacher-developed system.

Several issues related to scoring have emerged since implementation of alternate assessment was initiated on a broader scale. Among them is the question, who scores the assessment? In some states, teachers score their own students' work based on guidelines established in advance. States such as West Virginia use individuals who are involved in

a student's educational plan to evaluate their abilities. Virginia, for example, uses an outside contractor to score all collections of evidence. The greater issue, however, is that states base scoring of evidence on rigorous procedures that follow professionally accepted standards and statistical soundness (Bechard, 2001; Bolt et al., 2002; Hill, 2001; Quenemoen et al., 2002).

As the majority of states have conducted alternate assessment for less than four years, the reporting process may be among the areas receiving the greatest attention (R. Quenemoen, personal communication February 7, 2003). While states decide how to report alternate assessment results, they must simultaneously consider how results will be used as well as their impact on practice (Bechard, 2001; Hagan & Slocum, 2002; Quenemoen et al., 2002; Quenemoen & Thurlow, 2002; Thompson & Thurlow, 2000).

Principle 4: Accountability

Thurlow and colleagues (2001) defined this principle as follows: "The assessment performance of students with disabilities has the same impact on the final accountability index as the performance of other students, regardless of how the students participate in the assessment system" (Overview section). This principle is evident when (a) the performance data for all students have the same impact in accountability indices, (b) incentives exist for including all students in the accountability system, and (c) systems are held immediately accountable for all student performance although an appeals process exists as well as a mechanism for phasing in students previously denied access.

Now that development and implementation of alternate assessment are complete, states are focusing greater attention on reporting results. IDEA 97 requires that states report alternate assessment results in the same manner and frequency as they report

results on the general accountability system. This has posed a challenge for most states because alternate assessment dimensions may not be linked to a state's general education standards. Additionally, different proficiency levels and scoring systems may be used for the two types of assessments, thus creating a technical challenge. Alternate assessment has been in place for four years, yet states are still determining how to include alternate assessment results with those of the general assessment in their accountability system. State leaders who are confident that they have undergone rigorous standard setting feel somewhat confident that their alternate assessment results can be included in their accountability index (Quenemoen et al., 2002). Thurlow and colleagues (2001) contend that the mandate to integrate all performance results will put pressure on the integrity of the purpose of alternate assessment as well as its design.

Most states have undergone some form of standard setting to develop their proficiency levels. Alternate assessment, however, is still relatively new and standards continue to evolve. Roeber (2002) summarized six techniques, previously described by Cizek, that have been applied to standard setting for alternate assessment. They include (a) reasoned judgment, (b) contrasting groups, (c) modified Angoff (in which raters estimate and sum the percentage of students expected to pass each test item), (d) bookmarking or item mapping, (e) body of work, and (f) judgmental policy capturing. Each technique used should be based on the approach used for alternate assessment. Olson, Mead, and Payne (2002) hold that "...states must ensure that (a) assessments are aligned to content standards, and (b) performance standards have been set to determine the proficiency levels assigned to specific scores" (p. 1).

Regardless of the standard-setting approach employed, states must reevaluate

their scoring system if they begin to see large numbers of students participating in alternate assessment receiving high scores (Roeber, 2002). Thus, validity and reliability checks will likely play a significant role in the inclusion of all results as called for by NCLB in order to monitor adequate yearly progress. The ECS monitors state regulations, policy, guidelines, and so on, on nine requirements related to accountability under NCLB. To date, while only nine states and the District of Columbia have not identified how they will determine annual adequate yearly progress, almost half of the states lack evidence that all subgroups will be included in the accountability system (ECS, n.d.).

Principle 5: Improvement

This principle is evidenced when: “There is improvement of both the assessment and the accountability system over time, through the process of formal monitoring, ongoing evaluation and systematic training in the context of emerging research and best practice” (Thurlow et al., 2001, Overview section). Characteristics of this core principle include (a) all decisions are collected, compiled, and reported and the data is used to improve the assessment program; (b) consequences of student assessment decisions are collected and reported, and this information is used to improve accountability; (c) training is provided to various audiences secondary to monitoring and evaluation results; and (d) preservice and/or inservice training is provided for IEP teams and other personnel involved in the assessment process.

Early evidence suggests that positive consequences of including all students in assessment systems are beginning to emerge. For example, Rhode Island, Colorado, and Kentucky report 95-99% participation of students with disabilities in the general assessment (The Consortium on Inclusive Schooling Practices, 2001). Both the number

of students taking these exams and the performance of these students is rising. Including students with disabilities in assessment and accountability systems ensures that all students have access to the general curriculum. Quenemoen, Lehr et al. (2001) offered three broad recommendations for states to improve both assessment and accountability systems. They include:

1. Use a data-based continuous improvement approach to monitor systems
2. Identify all key stakeholders and maintain communication during implementation and change.
3. Keep standards high and maintain a clear focus.

Quenemoen and colleagues (2002) found that following the first mandated year of alternate assessment, states began revising their scoring processes, rubrics, and proficiency descriptors. Many provided additional training to teachers, parents, and other IEP team members. Some states are further along in the monitoring and evaluation of their assessment program and are critically looking at the reliability and validity of their assessments and correlations between assessment results and instruction. IDEA 97 supports the fact that systematic training is essential to the continuous improvement of inclusive assessment and accountability systems, recognizing the importance of both preservice and inservice education for individuals working with this population (Kleinert et al., 2000). Families, educational leaders, legislators, and school personnel must continuously monitor not only the extent to which students with significant disabilities are included but also the extent to which assessment results are reported and included in accountability systems and lead to improved student achievement (Kleinert et al., 2000; Quenemoen et al., 2002; Ysseldyke & Olsen, 1999).

Principle 6: Inclusive School Reform

In order to demonstrate evidence of this principle the following must exist: “Every policy and practice reflects the belief that *all students* [sic] must be included in state and district assessment and accountability systems” (Thurlow et al., Overview section). When assessment programs show alignment with this core principle, there is evidence of broad support from state legislative bodies and professional organizations for inclusion of all students in reform efforts. Additionally, all students are included in every component of the assessment and accountability system (e.g., reporting, determination of accountability measures, and use of data for improvement). There is also evidence that states are collaborating with various stakeholders as they design and evaluate their assessment and accountability systems.

The issue of including all students was implied in standards-based reform efforts but has often been absent in practice. Thurlow, Ysseldyke, Gutman, and Geenan (1998) found that even when state standards documents reported that all students were expected to achieve the standards, many of them did not identify students with disabilities. They also found that in many states, individuals who work closely with students with disabilities are not part of the standards development process. States that have not already done so must look closely at the purpose of their assessments and their beliefs and attitudes about what students with disabilities are able to learn (Consortium, 2001; Quenemoen, Thompson et al., 2001; Thurlow, Elliott et al., 1998).

Although the core principles espoused by Thurlow and colleagues (2001) are based on research from the past 10 years, they are the areas most evident in current literature on alternate assessment. This can be observed in the sample of the literature

addressing the core principles that follows in Table 1.

Table 1

Literature Addressing the Core Principles

Core Principles (Thurlow, Quenemoen, Thompson, & Lehr, 2001)	Inclusion of all students in assessment systems	Decisions based on clearly articulated participation guidelines	Scoring and public reporting	Inclusion of all results in the accountability system	Improvement through monitoring, evaluation, and training	Policies and practices that reflect the belief that all students should be included in assessment systems
Almond, Quenemoen, Olsen, & Thurlow, 2000		●				
Bechard, 2001			●	●		
Bolt, Krentz, & Thurlow, 2002		●	●	●		
Browder, Ahlgrim-Delzell, Flowers, Karvonen, Spooner, & Algozzine, 2002	●			●		
Browder, Flowers, Ahlgrim- Delzell, Karvonen, Spooner, & Algozzine, 2002					●	
Browder, Spooner, Algozzine, Ahlgrim-Delzell, Flowers, & Karvonen, 2003	●	●	●	●	●	●
Consortium on Inclusive Schooling Practices, 2001	●	●	●	●	●	●
Hager & Slocum, 2002	●	●	●			
Hill, 2001			●	●		
Kleinert, Haigh, Kearns, & Kennedy, 2000			●	●	●	
Krentz, Thurlow, & Callender, 2000				●		
La Marca, 2001			●	●		
Linn, 2001				●		
Linn, Baker, & Herman, 2002				●		

Core Principles (Thurlow, Quenemoen, Thompson, & Lehr, 2001)	Inclusion of all students in assessment systems	Decisions based on clearly articulated participation guidelines	Scoring and public reporting	Inclusion of all results in the accountability system	Improvement through monitoring, evaluation, and training	Policies and practices that reflect the belief that all students should be included in assessment systems
Olson, Mead, & Payne, 2002			●			
Quenemoen, Lehr, Thurlow, & Massanari, 2001	●			●	●	●
Quenemoen, Rigney, & Thurlow, 2002			●	●	●	
Quenemoen, Thompson, & Thurlow, 2003			●			
Quenemoen, Thompson, Thurlow, & Olsen, 1999		●				
Quenemoen & Thurlow, 2002			●	●		
Roeber, 2002			●			
Thompson, Quenemoen, Thurlow, & Ysseldyke, 2001	●	●	●	●	●	
Thurlow, Olsen, Elliott, Ysseldyke, Erickson, & Ahearn, 1996	●	●		●		
Thurlow, Ysseldyke, Gutman, & Geenan, 1998	●					●
Turner, Baldwin, Kleinert, & Kearns, 2000		●	●			
Warlick & Olsen, 1999		●				
Wiener, 2002			●			
Ysseldyke & Olsen, 1999		●			●	

Summary of Literature Review

Studies on state progress beyond implementation are just beginning to emerge (Browder, Ahlgrim-Delzell et al., 2002; Browder, Flowers et al., 2002; Wiener, 2002).

According to ECS (n.d.), generally, states are on the right track toward reaching the legislative mandates of NCLB to include all students in their assessment system, but few have made progress to including their performance in accountability indices. The web sites of such organizations as the ECS, the National Conference of State Legislatures, and the Council of Chief State School Officers provide various perspectives on the increased attention on accountability (Bolt et al., 2002).

As states work toward inclusive assessment and accountability systems, they must examine how students with severe to profound disabilities are included (Consortium, 2001; Elliott et al., 2000; Erickson et al., 1998; Hager & Slocum, 2002; Quenemoen, Thompson et al. 2001; Thompson et al., 2001; Thurlow et al., 1996, Thurlow, Ysseldyke et al., 1998). To do so, states must first examine their underlying assumptions about the value of assessing and including this population (Quenemoen, Thompson et al., 2001; Thurlow et al., 2001). States have initiated programs to comply with IDEA mandates while simultaneously wrestling with how this information will be reported, be used to make funding and program decisions, impact policy and, ultimately, result in increased student performance.

Decisions regarding who participates in alternate assessment will be more closely scrutinized under the mandates of NCLB (ECS, Hager & Slocum, 2002; NCLB, 2002; Paige, 2002). Additionally, as assessment programs and results become more visible to the public, reliable and valid instruments, decisions, and interpretations will become more crucial (Bechard, 2001; Bolt et al., 2002; Hill, 2001; La Marca, 2002; Linn, 2001; Quenemoen & Thurlow, 2002; Quenemoen et al., 2002).

Further, as states begin to evaluate current programs, they must simultaneously

use this data to improve them and provide training to key stakeholders (Browder, Flowers et al., 2002; Consortium, 2001; Kleinert et al., 2000; Quenemoen, Lehr et al., 2001, Quenemoen et al., 2002; Thompson et al., 2001; Ysseldyke & Olsen, 1999). Chapter 3 offers details regarding the methods used in this study to examine six states' alternate assessment policy. The study's design makes possible the collection of data that will provide insight into states' progress in meeting current legislative mandates for assessing students with the most significant disabilities.

CHAPTER III - Methods

All states are reported to have responded to the mandate to implement alternate assessment (Thompson & Thurlow, 2001). Now they must include and report alternate assessment results together with those of all other assessments using a single accountability system. Further, states must continue to monitor the status of alternate assessment implementation, particularly within the context of current policy mandates. Aside from status reports on the implementation of alternate assessment, there is a dearth of published literature on state activities to determine whether their programs are congruent with the core principles of inclusive assessment and accountability systems. Researchers believe that such alignment is paramount if these measures are to be reported in the single accountability system called for by NCLB (Browder et al., 2003; NCEO, 2003; Quenemoen et al., 2003).

This chapter describes the research methods used in the present study and provides the following: (a) a restatement of the research questions, (b) a rationale for the use of a mixed design incorporating content analysis and grounded theory, (c) the instrumentation and protocol for each phase of the study, (d) a description of the data analysis selected, and (e) a discussion of the ethical safeguards.

State data and documentation exist to provide a means of examining the status of alternate assessment policy and the existence of the core principles and characteristics described in the literature. This study analyzed data collected from various sources to

answer the following research questions:

1. Are the core principles and characteristics of inclusive assessment and accountability systems reported in the literature present in legislation, policies, and procedures that govern state alternate assessment programs?
2. What similarities exist across state alternate assessment policy
3. What differences are evident in state alternate assessment policy?
4. What additional themes emerge from analysis of the state legislative, policy, and procedure documents that drive alternate assessment programs?

Design

The study employed a mixed design using content analysis and grounded theory. This allowed the researcher to make inferences regarding state alternate assessment policy based on evidence of the core principles in assessment and accountability documents and to identify emerging themes. Content analysis “aims to help improve the quality of inferences made by analysis of communication” (Carney, 1972, p. 26). The process of making inferences from the data to specific aspects of their context, based on knowledge of the system, facilitates use of the data to provide information on the subject or issue being examined (Krippendorff, 2004). Additionally, a grounded theory approach as defined by Strauss and Corbin (1998) was used to systematically compare and analyze the data. This approach allowed the researcher to begin with a topic and let the theory emerge from the data during the research process.

The research was accomplished in three phases. In the first phase, the NCEO guide, *A Self-Study Guide to Implementation of Inclusive Assessment and Accountability Systems: A Best Practice Approach* (Quenemoen, Thompson et al., 2001) provided the

categories that were used in the content analysis. Using this guide, codes and indicators were developed along with the coding forms used to collect the data. Using a sample of state documents, the author and a second coder then tested the codes to establish their reliability before proceeding to the next phase.

Phase Two involved examining written communication on assessment and accountability from Kentucky, Maryland, North Carolina, Tennessee, Virginia, and West Virginia against the categories and their indicators established in Phase One. These states were selected because they reflected differences in their (a) special education demographics, (b) date of alternate assessment implementation, (c) linkage to funded higher education projects, and (d) history of inclusive assessment. This phase explored the presence and frequency of the characteristics of best practice reported in the literature.

The third phase of the study involved a qualitative content analysis using a grounded theory research approach. Specifically, the alternate assessment policies of the six states were examined for similarities and differences and to determine if additional themes exist beyond those cited in the literature as principles of best practice. Although states are bound by the same federal mandates regarding the assessment and inclusion of students with disabilities, implementation of the policy is subject to their interpretation yielding input from multiple “interpretive communities” (Yanow, 2000, p. 223). It was anticipated that such examination would offer insight into how states have interpreted the technical aspects of NCLB and IDEA 97 in constructing state policy documents governing alternate assessment.

Instrumentation and Protocol

Phase One

Content analysis is defined as “any technique for making inferences by objectively and systematically identifying specified characteristics of messages” (Carney, 1972, p. 25). The use of this technique not only enables the researcher to summarize the formal content of written material but also to describe the attitudes or perceptions of the author of the material as well as the intended and actual audience (Gall et al., 1996; U.S. General Accounting Office [GAO], 1996). In brief, this technique enables the researcher to make valid inferences from the data to particular aspects of their context and to justify these inferences in terms of what is known about the system being examined (Krippendorff, 2004; Weber, 1990).

Various recommendations exist for conducting a content analysis. The GAO (1996) identified seven steps in conducting a content analysis (a) deciding whether to use content analysis; (b) defining the variables; (c) selecting material for analysis; (d) defining the recording the units (i.e., by-document); (e) developing an analysis plan; (f) coding the textual material; and (g) analyzing the data. These steps, and recommendations by other authors, stress not only the importance of reliability and validity but also the importance of sample selection, determining the unit of analysis, creating a coding scheme, and data analysis (Carney, 1972; GAO, 1996; Krippendorff, 2004; Weber, 1990).

This study included the following steps: (a) identifying material for analysis (described under *Document Sample*), (b) determining the coding unit, (c) identifying the categories, (d) considering reliability issues, (e) considering validity issues, (f) analyzing

the data, and (g) ensuring ethical safeguards. With the exception of ethical safeguards, which will be addressed later, each of these steps will be described below.

Determining the coding unit. Defining the basic unit of text to be classified is one of the most important determinations in a content analysis (Weber, 1990). The basic unit is the portion of text to which a category label is applied (GAO, 1996). The most common options include the following:

1. **Words:** Words are clearly defined recording units with identifiable physical boundaries. Words are easily classified by computers and are a reliable option to use as a recording unit. Some computer software, however, may have difficulty distinguishing words with multiple meanings.
2. **Word Sense:** Word sense is a variation on words as the recording unit. Some computer programs are able to distinguish between multiple meanings of words and identify phrases that represent semantic units, which can be counted as if they were words.
3. **Sentences:** Sentences are an appropriate unit when the researcher is interested in words or phrases that occur closely together. Using sentences as a recording unit requires human coding, which can affect reliability.
4. **Paragraphs:** When limited human or computer resources are available, the researcher may code a paragraph. Caution should be exercised as paragraphs sometimes contain too many ideas for reliable assignment of text to single categories.
5. **Themes:** Themes are useful recording units because the boundary of the theme describes a single idea. Themes may present problems of reliability

because they require coder judgment, however.

6. Whole Text: Whole text is larger than a paragraph and consists of clearly defined boundaries. Whole text coding is highly unreliable.

(GAO, 1996; Weber, 1990)

The coding unit considered the most appropriate for this study was the theme because of the desire to capture information in context. Themes, however, require coder judgment and may present problems of reliability. Test coding was conducted to address this potential problem. Gall et al. (1996) recommend that asking a second person to apply sample text to the coding categories is useful to discovering problems inherent in the coding scheme. This is further described in the section on reliability and validity.

Identifying the categories. Emerging best practice suggests that by examining a state's assessment and accountability system against specific criteria, administrators can identify underlying assumptions and beliefs that drive decisions and minimize unintended consequences (Thurlow et al., 2001). The variables used in this study were derived from the NCEO Self-Study guide (Quenemoen, Thompson et al., 2001). The categories assigned to each variable (core principle) were the indicators identified by NCEO and ranged from one to five categories per variable.

A code was created for each category and defined in a coding manual to enable a second or subsequent coder to be trained and for use during the main study. This manual contains the Guidelines for Coding (list of codes, what they mean (criteria), indicators, examples of the code in text), Coding Instructions, and the Content Analysis Coding Form (see Appendixes A, B, and C).

Considering reliability issues. Reliability refers to the extent that any research design represents variation in real phenomena, rather than the circumstances of measurement, the hidden idiosyncrasies of the analyst, and the biases of a procedure (Krippendorff, 1980). The term encompasses at least three types of reliability that the researcher must consider when designing a content analysis: (a) stability, (b) reproducibility, and (c) accuracy (Krippendorff, 2004).

Stability refers to the extent that a process is unchanging over time. For example, stability can be determined when one analyst codes the same content more than once. Because it relies on a single coder, it is the weakest form of reliability. Reproducibility refers to the extent to which two or more coders produce the same results. This form may also be referred to as “inter-coder reliability” (Krippendorff, 2004, p. 215). High reproducibility is necessary in content analysis. The last type, accuracy, is the strongest form of reliability, and refers to the extent to which the categorization of text corresponds to a standard or norm (Krippendorff, 2004; Neuendorf, 2002; Weber, 1990).

To satisfy issues of reliability, this researcher sought high reproducibility. A sample of 10 documents was coded by the primary researcher and by a second coder to discover problems inherent in the coding procedure. Because coding tends to be faster and more accurate when the coder is very knowledgeable about a subject, test coding was completed by the researcher and a doctoral student who was knowledgeable about alternate assessment. Both individuals coded the same documents. The test coding process consisted of the following steps: (a) selection of a second coder knowledgeable about assessment and accountability; (b) training the second coder in the coding process; (c) test coding the sample of 10 documents with a goal of 80% consistency between

coders; and (d) if 80% consistency was obtained, stopping the test coding process, if not, test coding 10 additional documents. A percent agreement coefficient was used to determine scorer/rater reliability:

$$Pa_o = A/n$$

Where, Pa_o = proportion agreement; A = the number of agreements between coders; and n = the total number of items the two coders have coded (same as the maximum possible number of agreements the coders could obtain) (Neuendorf, 2002).

Considering validity issues. Validation must occur to ensure the credibility and sturdiness of the research findings in order to use them when developing theory or in making practical decisions (Miles & Huberman, 1994). This is particularly important if the results from the content analysis are intended to (a) have policy implications, (b) aid government and industry, (c) be used as evidence in court, or (d) affect individual human beings (Krippendorff, 1980).

The validity of the results is more powerful to the extent that other data, coding procedures, or classification schemes produce substantive conclusions. Carney (1972) suggested that in order to confirm the validity of a study's inferences, the researcher test them against materials not considered in the original design. Correspondence between the category and the concept it represents produces stronger validity. Thus, construct validity is the extent that a measure correlates with other measures of the same construct and does not correlate with dissimilar constructs (Weber, 1990).

In this study, the researcher first sought to establish credibility of the findings by determining if the conclusions drawn via qualitative and quantitative methods converged. The researcher further sought to determine if a correspondence between the indicators of

best practice evident in the literature and the data existed to strengthen the plausibility of the findings. Additionally, validity was sought by examining the data's congruence with core principles established in prior research. Although the findings from this investigation may be limited in their transferability due to the delimitations imposed, the process is described in sufficient detail to permit its application to other states.

Phase Two

Data sources. Phase Two involved examining documents on assessment and accountability from KY, MD, NC, TN, WV, and VA for the categories and indicators established in Phase One. Initially, the researcher considered including all states in the 4th Circuit since they would be similarly affected by case law; however, because of the relative newness of alternate assessment and recent implementation of NCLB, such effects have not yet occurred. Consequently, states were selected instead that reflected a range across four characteristics making them a more representative sample. First, the states either had a longstanding history with alternate assessment or had recent implementation. Further, states demonstrated a history of inclusive assessment or possessed demographics shared by more than half of the states. Finally, states had ties to funded higher education projects with faculty and staff conducting the research on emerging alternate assessment practice.

Of the states in the sample, Kentucky and Maryland had alternate assessment programs in place before the IDEA mandate while the others met the implementation deadline of July 1, 2000. Three of the states, Kentucky, Maryland, and North Carolina, are currently linked to funded university projects that provide technical assistance and training as well as conduct research on assessment practices and results. The

Interdisciplinary Human Development Institute at the University of Kentucky hosts the Kentucky Alternate Portfolio (KAP) Project under the direction of Jacque Farmer Kearns (KAP Project, n.d.). The Institute for the Study of Exceptional Children and Youth at the University of Maryland (UM) is one of three organizations collaborating to form the Educational Policy Reform Research Institute (EPRRI). This institute is a federally funded program aimed at increasing knowledge and understanding of ways that students with disabilities can be included in educational accountability measures (Educational Policy Reform Research Institute, n.d.). In addition, investigators from the University of North Carolina at Charlotte generate much of the literature related to curriculum and instruction for students participating in alternate assessments via three funded projects. They include (a) the Charlotte Alternate Assessment Model Project, (b) the Evaluation of Emerging Alternate Assessment Practices Project, and (c) the General Curriculum Access for Students with Significant Cognitive Disabilities Research Project (“The Impact of Alternate Assessment,” 2003).

The number of children ages 6 to 21 served under IDEA, Part B, in the 2000-2001 school year for the states included in this study ranged from 44,888 (WV) to 155,706 (NC), with an average of 122,343. Further 56% and 64% of the 50 states and the District of Columbia fall within this range for students served between the ages of 6-21 and 3-21, respectively (U.S. Department of Education, 2002).

These six states were also included to determine if their proximity to one another had any influence on their interpretation of the mandates and resulting state policies. Both KY and TN, for example, had early involvement in the development of the VA alternate assessment program and other such involvement is common across states. It was

recognized that this might pose limitations on making generalizations about the findings.

Document sample. The Consortium on Inclusive Schooling Practices (2001) offers a list of potential source documents to determine policy supports for inclusive schools. They include state guidance on assessment; training materials, rules, and guidelines; state alternate assessment documents and guidelines; state accountability policy; state guidance on accountability systems; state assessment policy; state accountability plans; state assessment legislation; state data forms; and state professional development policy and records. These items represent examples of written communication that contain key ideas and use language to convey messages.

The sample of written communication for the study consisted of the body of state documents that govern the areas of, report the status of, or provide guidance regarding assessment, accountability, and special education. In this study, documents that represent state legislation, policy and procedure, reports, and manuals were analyzed. Examples of documents from categories that were examined are found in Table 2.

Table 2

Examples of State Documents Included in the Content Analysis

Legislation	Policy Documents	Manuals	Reports
State Statutes	Consolidated State Application	State Assessment Program	BOE Annual Report
Administrative Code	Accountability Workbook	Administration Manual	Findings and Recommendation of the Advisory Committee on Testing and Accountability
	Guidelines for Participation of Students with Disabilities	Alternate Assessment Program Implementation Manual	

Other items such as superintendents' memos and committee meeting minutes were used to identify documents for the sample but were not included in the analysis.

Some of the documents in the sample are published or revised annually.

Therefore, only documents published from 2000-2003 were considered unless their timeframe was broad and included these years. First, a search of the state's Department of Education (DOE) website was conducted for documents meeting the above criteria. The search was completed using the following search terms: *alternate assessment, standards and special education, disability and assessment, NCLB and subgroups, and accountability*. Next, the state's Board of Education (BOE) website was searched for documents related to assessment, special education, and accountability. Finally the state's legislative website was searched for codes and statutes dealing with education.

After obtaining this initial set of documents, the researcher reviewed them to identify additional documents relevant to the study. The researcher then emailed the alternate assessment specialist at each state department of education a list of the documents retrieved to determine the completeness of the data sources and to solicit any additional documents that may have been overlooked. This step also provided verification of the data sources. The documents included in the content analysis are listed by state in Appendix D.

To manage the potentially large volume of data, after recording the source information and type for each document on the Coding Form, the researcher affixed a color-coded tab to the document and placed it in a colored expandable file. This was done to facilitate easy identification of documents across states and type. Two additional copies of the documents were photocopied on colored paper using a different color for

each state to use in the subsequent analyses. Upon the completion of these steps, each research question was answered using the data collection and analysis methods that follow.

Research Question 1. Content analysis was used to answer Research Question 1: “Are the core principles and characteristics of inclusive assessment and accountability systems reported in the literature present in legislation policies, and procedures that govern state alternate assessment programs?” Content analysis may focus solely on the presence of a variable, how frequently it appears, its intensity, or the space or time devoted to it in a document (GAO, 1996; Krippendorff, 2004; Neuendorf, 2002). This phase of the study examined the presence and frequency of the categories and indicators developed in Phase One.

The content analysis was conducted during the winter of 2003-2004. Documents from KY, MD, NC, TN, VA, and WV were analyzed. The researcher first examined the documents for evidence of the indicators of best practice identified by NCEO using their six core principles and accompanying Alternate Assessment Self Study Checklist (Quenemoen, Thompson et al., 2001). The checklist contains a set of 16 indicators that the authors contend represent characteristics (categories) of inclusive assessment and accountability systems. A summary of the categories and research developed codes and indicators are presented in Appendix E.

The procedure developed for test coding was used during this phase. Using the Content Analysis Coding Form that was initiated during the retrieval process and the code manual also developed in Phase One, the researcher manually examined each document for evidence of the 16 categories. Specifically, the text was marked each time

the indicator was observed and the information was recorded on the coding form. For each of the 16 indicators on the form, the researcher noted whether there was “evidence,” “no evidence,” or evidence that the indicator was “in progress.” Also, the page number(s) for each occurrence of key text was recorded. Key text was considered the portion of text that included an indicator. After determining the existence and prevalence of the categories for each state, this information was analyzed in the next phase across the states.

Phase Three

Research Questions 2 and 3. There are numerous ways to generate meaning from research data. Some of the more descriptive tactics offered by Miles and Huberman (1994) include noting patterns and themes, seeing plausibility, clustering, and making contrasts/comparisons. However, description is only the first step in conducting research in what Strauss and Corbin (1998) define as grounded theory approach. Description serves as the basis for the additional data gathering requisite for theory development and explanation.

A descriptive account of the similarities and differences in states’ alternate assessment policy is presented later in the findings using the information gleaned from the analysis of data generated in Research Question 1. To complete this analysis, the text coded for each indicator was “cut and pasted ” on index cards, resulting in 16 piles of text for each state. Then, the information was examined across states to establish characteristics that were similar and those that were different for each category.

Research Question 4. Frequency counts assume that the most frequently appearing categories reflect the greatest concerns and reveal differences between

documents (Weber, 1990). It was anticipated that the categories identified in the literature were not the only ones that existed in state documents. To explore this assumption, a second level of qualitative analysis was used to answer Research Question 4, “What additional themes emerge from analysis of the state legislative, policy, and procedure documents that drive alternate assessment programs?”

Strauss and Corbin (1998) described a process for completing an analysis that involves open coding, axial coding, and selective coding. Such an approach may be used to uncover meaning the research draws from the data. Open, axial, and selective coding were used; the process used in this study is explained in further detail below.

Data Analysis

Research Question 1

After all the documents had been coded, the coding forms were assembled, separated by state. The number of occurrences for each indicator was then summed and the total number of coded occurrences totaled. Then, the total number of indicators for each category divided by the total number of codes across all documents was calculated to rank the categories from most prevalent to least prevalent.

Using descriptors that match the language in the code manual, all instances were counted in which a category was coded in a document. These categories represent what the literature identifies as components of the core principles, (a) all students, (b) decisions, (c) reporting, (d) accountability, (e) improvement, and (f) inclusive school reform across all settings. A table was used to summarize the number and percentage of occurrences for each category. Then, the principles were ranked from the greatest number of codes to the least observed. This was done for each state. The prevalence of

each indicator within the six categories was then examined for all states and patterns were noted across states.

Research Questions 2 and 3

The coded text was examined across states for similarities in the six categories (core principles) and for differences in each category. First, similarities in how the categories ranked across the states were examined to look for patterns. Then, exceptions to these patterns were explored, and examples of trends in a state's documents or language in the text were identified. This was followed by an examination of how the 16 indicators were addressed in each state.

Research Question 4

A second level of analysis was conducted to determine additional categories that emerged as constructs apart from the indicators of best practice identified in the literature. To complete this analysis, open coding and axial coding described by Strauss and Corbin (1998) was used. The goal of this analysis was to identify categories within the data, identify the characteristics of those categories, and establish their relationship, if any, to the core principles espoused in the literature.

A clean, colored copy of each document was used for this stage of the research. The documents were re-read, and all sections addressing or related to alternate assessment were cut out and glued to large sheets of chart paper to facilitate coding. Although easier to work with, because of the large amount of text that would be involved, it was not practical to use index cards for this task. A process of open coding, axial, and selective coding, as described below, was completed.

Open coding. In open coding, data are broken down into discrete units, examined closely, and compared for similarities and differences. Initially, each segment of data conveying a single idea was examined and a concept listed on a blank sheet of chart paper. Memos regarding any question or thought dealing with a segment of text were made on the right side of the page while key concepts or codes were written down the left side. Concepts representing the central ideas in the data were drawn from the document segments. Initially, ideas that were considered similar were grouped together. According to Strauss and Corbin (1998), this initial conceptualization permits the researcher to “open up the text” (p. 113), although dozens of concepts may be initially identified.

Upon reviewing the preliminary codes, it was noted that some of the concepts seemed to fit multiple categories or could be grouped into an even broader category. The data were then re-examined and new labels generated in some cases, resulting in fewer concepts. Next, a more detailed examination of the data was performed to identify characteristics and any variations of the characteristic of each category. This process was facilitated by the memos made while the preliminary codes were generated.

Characteristics, identified by Strauss and Corbin as properties included “the general or specific characteristics or attributes of a category ” (p. 117). Variations of a property “represent the location of a property along a continuum or range (p. 117). For example, a property that distinguishes a “novice” from an “advanced” computer user is “frequency” or number of hours per day of computer use. Subcategories would be considered in the next level of coding, axial coding.

Axial coding. Axial coding is the “act of relating categories to subcategories along the lines of their properties and dimensions. It looks at how categories crosscut and link”

(Strauss & Corbin, 1998, p. 124). A category represents a phenomenon; in other words, it is a problem, issue, or occurrence that is significant to those involved. A subcategory is also a category; however, it provides further explanation regarding who, what, when, where, and how. Strauss and Corbin offered four basic tasks in the axial coding procedure:

1. Laying out the properties of a category and their dimensions, a task that begins during open coding
2. Identifying the variety of conditions, action/interactions, and consequences associated with a phenomenon
3. Relating a category to its subcategories through statements denoting how they are related to each other
4. Looking for clues in the data that denote how major categories might relate to each other. (p. 126)

Selective coding. Selective coding involves integrating and refining categories.

Strauss and Corbin (1998) suggested that although data at this stage no longer represent a single case (i.e., state), categories are derived by comparing data from each case; therefore, they should have relevance for and be applicable to all cases in the study (p. 145). After identifying an initial set of categories and examining the relationships among the categories, a central phenomenon is sought. This phenomenon, according to Strauss and Corbin, must be one that the other categories can be related to. Additionally, there should be numerous indicators pointing to the phenomenon; giving it an ability to pull all of the other categories together to develop (what Strauss and Corbin refer to as “explanatory power,” [p. 147]). This process required repeated examination of the

categories and recategorization of data after re-reading all memos.

After re-examining all of the data, major categories were related to the central phenomenon by offering an explanation of the relationships. To accomplish this, chart paper and sticky notes were used to create diagrams of the relationships while writing notes of explanation. Chapter IV contains a table of the major categories and subcategories that emerged during the coding. Finally, Strauss and Corbin (1998) pointed out that it is more important to address the interrelatedness of the concepts, not just list emerging themes. Both the themes that emerged during the coding process and their relationship to one another are discussed in the next chapter. Table 3 provides a summary of the data collection, data sources, and data analysis used to answer each research question.

Table 3

Specifications for Data Analysis

Research Question	Data Collection Method	Data Source	Data Analysis
1. Are the core principles and characteristics of inclusive assessment and accountability systems reported in the literature present in legislation, policies, and procedures that govern state alternate assessment programs?	Content analysis	State assessment & accountability documents	Quantitative; frequency counts
2. What similarities exist across state alternate assessment policy?	Content analysis	State assessment & accountability documents	Qualitative; descriptive

Research Question	Data Collection Method	Data Source	Data Analysis
3. What differences are evident in state alternate assessment policy?	Content analysis;	State assessment & accountability documents	Qualitative; descriptive
4. What additional themes emerge from analysis of the state legislative, policy, and procedure documents that drive alternate assessment programs?	Grounded theory	State assessment & accountability documents	Qualitative; open, axial, and selective coding

Ethical Safeguards

The design of this study was exploratory and involved the use of existing documents. It was not obtrusive, nor did it require any interventions or treatments with human subjects. Consistent with federal, state, and university policy, the research did not require Protection of Human Subjects Committee approval.

CHAPTER IV - Data Analysis

The literature on accountability and assessment suggests that states must include all students, base participation decisions on students' ability to participate in the general assessment, report all results, include those results in accountability indexes, and use the results to improve instruction and student performance in order to realize inclusive reform (Quenemoen, Thompson et al., 2001). The purpose of this study was to determine if the core principles cited in the literature were evident in six states' alternate assessment policy and procedure documents and to identify if additional themes beyond those found in the literature emerged from these documents.

The states examined included Kentucky, Maryland, North Carolina, Tennessee, Virginia, and West Virginia. A mixed design using content analysis and grounded theory was employed to gain insight into states' alternate assessment policy. The study was completed in three phases. In Phase One, the NCEO guide, *A Self-Study Guide to Implementation of Inclusive Assessment and Accountability Systems: A Best Practice Approach* (Quenemoen, Thompson et al., 2001) provided the categories that were used. Using this guide, codes and indicators were subsequently developed along with the coding forms used to collect the data. The codes were then tested to establish their reliability before proceeding to the next phase.

Phase Two involved examining written communication on assessment and accountability from the six states against the categories and their indicators established in

Phase One. Documents were coded for evidence of the 16 indicators reflecting the core principles. In Phase Three, a descriptive account of the similarities and differences in states' alternate assessment policy was completed using the information gleaned from the analysis of data generated in Research Question 1. After this cross-state analysis, clean copies of the documents were used and a process of open and axial coding was completed to uncover additional themes beyond those cited in the literature as core principles. This chapter presents an analysis of each phase of the study. The subsequent discussion includes (a) a summary of the development of the code manual, coding forms, and test coding process; (b) a discussion of the core principles evident in state alternate assessment legislation, policy, and procedures; and (c) the identification of major themes that emerged during the grounded theory process.

Phase One

The design of the content analysis used in Phase Two was based on pre-existing categories identified in the literature. Although the categories and their characteristics were identified a priori (Quenemoen, Thompson et al., 2001), it was necessary to develop a code manual that included a list of codes; what they meant (criteria); indicators; and examples of the code in text to facilitate reliable coding by multiple coders. Additionally, Coding Instructions, and the Content Analysis Coding Form were also created.

Test coding was conducted in August 2003 by the researcher and a doctoral student familiar with alternate assessment to address the potential problem of coder judgment. A percent agreement coefficient was used to determine scorer/rater reliability. Although a proportion agreement of 92% was achieved, revisions were made to several of the codes before beginning Phase Two because of several instances of coder

disagreement and the need to clarify some of the indicators. A summary of the test coding results is located in Appendix F.

Phase Two

Research Question 1

Content analysis was used to answer Research Question 1: “Are the core principles and characteristics of inclusive assessment and accountability systems reported in the literature present in legislation, policies, and procedures that govern state alternate assessment programs?” By reviewing documents from the six states, the researcher was able to identify which core principles existed in individual states’ policy documents. (A list of the documents included in the analysis may be found in Appendix D.) A numerical and descriptive summary of the existence and prevalence of the principles and their indicators per state follows. An attempt was made to obtain like documents for each state when such existed. However, because of differences in the states’ assessment and accountability systems this was not always possible.

While it was necessary to obtain the number of occurrences for each indicator, using this information alone would yield inaccurate data because of the variation in the number and types of documents across the six states. For this reason, the total was converted to a percent for each core principle. This allowed the researcher to rank and make some comparisons across the states regarding the most to least prevalent principle. Table 4 provides a summary of these data.

Table 4

Core Principles Noted in State Policy Documents

	KY # (%)	MD # (%)	NC # (%)	TN # (%)	VA # (%)	WV # (%)
All Students (AL)	15 (20)	11 (30)	28 (31)	12 (32)	27 (23)	11(18)
Decisions (DE)	16 (21)	9 (25)	14 (16)	9 (24)	47 (40)	14 (22)
Reporting (RE)	16 (21)	6 (17)	27 (30)	5 (14)	18 (15)	13 (21)
Accountability (AC)	6 (8)	2 (6)	7 (8)	2 (5)	2 (2)	6 (10)
Improvement (IM)	6 (8)	5 (14)	5 (6)	1 (3)	11 (9)	6 (10)
Inclusive Reform (IN)	16 (21)	4 (11)	8 (9)	8 (22)	13 (11)	11 (18)

As illustrated, every state revealed evidence of all of the core principles; however, the breakdown by category is revealing. For the first principle, “All Students,” while every state evidenced the principle (18%-20%), the majority of the evidence supported that alternate assessments are aligned or linked with state content standards. For the second principle, “Decisions;” again, the evidence ranked high for all states (16%-40%). There was a great deal of variation within this principle, however. Virginia, West Virginia, North Carolina, and Tennessee had the most evidence addressing the fact that the general assessment is the starting point for determining participation in the alternate assessment. Virginia documents included this point 29 times compared to the next closest state (NC) with 10. The foci of this indicator in Kentucky were twofold: the percent of students expected to participate and providing a clear explanation of the process for compiling data. Maryland documents focused more so on describing the process for compiling data.

The third principle, “Reporting,” contained the most categories; however, it also contained the categories coded the fewest number of times. Specifically, all six states lacked evidence in the documents that (a) state and district reports included the number and percent of students participating in the alternate assessment, and (b) rubrics are developed and studied for face validity. The number of instances coded ranged from 0-2, with North Carolina coded twice for both categories. Three states (NC, KY, and MD) were coded highest for evidencing a detailed approach for test administration, including reliability checks. In Virginia and West Virginia, the data focused on the inclusion of alternate assessment scores in state and local reports.

States have reported their intentions regarding the integration of alternate assessment scores in accountability indexes. Yet, it may be too early to see evidence of this, beyond reporting, in state documents. While the principle “Accountability” was coded for all states, this theme was found in all states’ NCLB Accountability Workbook submissions. “Improvement” was coded the fewest number of times for five of the six states, accounting for 2%-10% of the tags. For the last principle, “Inclusive School Reform,” results varied widely. In Kentucky, Tennessee, and West Virginia inclusive reform categories accounted for 19%-21% of the codes. Conversely, in Maryland, North Carolina, and Virginia the same principle demonstrated only 9%-11% of the codes. Table 5 illustrates how the principles ranked for each state.

Table 5

Indicators Ranked in Order of Prevalence by State

	KY		MD		NC		TN		VA		WV
	%		%		%		%		%		%
DE	21	AL	30	AL	31	AL	32	DE	40	DE	24
RE	21	DE	25	RE	30	DE	24	AL	23	AL	19
IN	21	RE	17	DE	16	IN	22	RE	15	IN	19
AL	20	IM	14	IN	9	RE	14	IN	11	RE	17
AC	8	IN	11	AC	8	AC	5	IM	9	AC	12
IM	8	AC	6	IM	6	IM	3	AC	2	IM	10

Note. DE = Decisions; RE = Reporting; IN = Inclusive school reform; AL = All students; AC = Accountability; IM = Improvement.

For all states, the largest percentage of evidence was found in policy and procedure documents and manuals (58%-76%), with little or no evidence of the core principles found in reports (0%-5%). For all states, the greatest concentration of information was found in assessment manuals. “All Students” or “Decisions” ranked highest in five of six states, while “Accountability” ranked fifth or sixth across all states. Except for Maryland, “Monitoring” was also among the bottom two categories for all states. Surprisingly, “Inclusive Reform Efforts” ranked third for three states. No two states were alike in the ranking of the principles.

Phase Three

Research Questions 2 and 3

Although approximately a quarter or more of the evidence for Research Questions 1 and 2 came from legislation for all states, the level of detail varied widely. For

example, Maryland legislation included very general language such as "... policy shall provide for annual training of appropriate personnel on the local test administration policy and procedures" and "For any test instrument authorized for use in a State mandated testing and measurement program, the Board shall recommend procedures and standards" The language in Virginia's documents was very similar. North Carolina's legislative documents, on the other hand, were very detailed. Over 20% of the coded material from North Carolina came from the Board of Education Policy Manual or the General Statutes. The State Board of Education Policy Manual contained such information as North Carolina Alternate Assessment Portfolio Scoring and Reporting Procedures and Accountability Standards for Students with Disabilities. Much of the documentation in Kentucky legislation reflected the existence of a single statewide accountability system, with reference to alternate assessment mainly as one component of the state assessment program. Rarely did separate documents exist for Kentucky on alternate assessment, except as attachments or appendices. Instead, specific information related to the assessment was included in the CATS Interpretive Guide or the District Assessment Coordinator Implementation Manual (DAC Guide) (Kentucky Department of Education, 2002, 2003).

While Tennessee had the most number of documents coded, it had among the fewest number of coded segments of text. The majority of the labels were applied to ideas in the Special Education Manual; however, the text dealt primarily with procedural aspects of the assessment such as guidelines for participation, collecting data, and the scoring process. Four of the indicators were not evident: (a) the inclusion of the number and percent of students participating in alternate assessments in state and district reports,

(b) the development of rubrics that are studied for validity, (c) the existence of training for key personnel, and (d) the integration of alternate assessment practices. The principles of “All Students,” “Decisions,” and “Inclusive Reform” were the three that ranked highest for both Tennessee and West Virginia, with both states demonstrating the same order of prevalence for the remaining three categories. Reporting was followed by “Accountability,” then “Improvement.”

The greatest number of documents coded was from Virginia. This may have been related to the fact that this state has separate manuals for its general, alternate, and substitute assessments. The alternate assessment program is referenced in the documents for both of the other assessments, while this was seen to a much lesser degree for the other states. Two of the categories were most prevalent in the Virginia document: the alignment of the alternate assessment with the general assessment (coded 21 times) and that participation decisions are based on student ability (coded 29 times). These two categories accounted for 42% of the coded segments and came from the assessment manuals. North Carolina was the only other state with as high a number (20) for the category related to the alignment of its alternate assessment. Of the state reports examined, Virginia’s Special Education State Improvement Plan Report provided the most evidence. Specifically, it addressed the number and percentage of students participating in the alternate assessment as well as performance data across content areas and performance levels.

Language specific to the West Virginia Alternate Assessment was prevalent across all types of documents - legislation, policy, manuals, and reports. In almost all instances, when the state’s general assessment, the WESTEST, was mentioned, so was

the alternate assessment. For example, the Board of Education Policy Manual cites in Policy 2340 §126-14-6 that “The West Virginia Department of Education shall disaggregate and report WESTEST and Alternate Assessment results by subgroups identified in the approved West Virginia Consolidated Application as required by NCLB.” The same language was included in the NCLB Accountability Workbook; “The WESTEST and WV Alternate Assessment results are the primary indicators on which AYP determinations are made for public schools and LEAs” (p. 44). Compared to the other states, West Virginia and Kentucky had the least amount of variation in the prevalence of the six principles ranging from 24% to 10% and 21% to 8%, respectively.

Research Question 4

Using clean colored copies of the documents coded in Phase Two, a continuous comparative process was used to uncover meaning in the research through the data (Strauss & Corbin, 1998). Using a constant comparative approach, the data were analyzed by coding; chunking; identifying categories, subcategories, and their characteristics and dimensions; and using memos and diagrams (Creswell, 1998; Miles & Huberman, 1994; Strauss & Corbin, 1998). The coding process continued until it was felt that the categories were “saturated,” that is, “no new information seemed to emerge during coding ...” (Strauss & Corbin, 1998, p. 136). The following themes emerged from the analysis: resources, accountability, improvement, inclusion, access, and reporting. A summary of the resulting categories, subcategories, properties, and dimensions is represented in Table 6. In addition, each theme will be briefly discussed.

Table 6

Results of Open and Axial Coding

Categories	Subcategories	Properties	Dimensions
Resources	Procedures	Guidelines Policies Information	General, detailed, step-by-step, timely, responsive
	Training/Assistance	Availability Content	Time, location, provider, recipient
Accountability	Results	All students	Format, use, reliability, validity
	Decisions	Eligibility Evidence	Responsibility
Improvement	Outcomes	Student Program System	Goals, targets, requirements, measurements, calculations
Inclusion	Assessment	Students with disabilities	Criteria, exceptions, exclusions
Access	Curriculum	General Education Functional	Standards, alignment, link, extension
	Instruction	Authentic Ongoing Appropriate	Real-life, age-appropriate, embedded, peers, setting
Reporting	Results	Participation	Number, percent
		Performance	Level, proficiency, content area

Resources. As Hord (cited in Meister, 2000) explained, change “must be viewed as a journey by individuals who have highly personal views and levels of understanding, which are evident in the different ways that they develop through a change endeavor” (p. 28). Research suggests that institutions go through predictable stages when an

innovation is introduced (Howley-Rowe & Leopold, 2000). Among these stages are operational concerns. The emphasis at this stage, which the authors suggested might last for two or more years, is on logistics, materials, schedules, and the equipment necessary to make the innovation work (Human Resource Development Press, 1995).

In the current study the greatest amount of text in state policy documents was found to center on this theme. Thus, procedural guidelines consumed most of the focus of state documents. This text was very similar across states, as most included statements such as “A complete portfolio will include the following items” or “Collections of Evidence will be gathered starting in October 2003 and continued through spring 2004” and “... eligible students will participate in ALT-MSA in grades 3-8 and 11.” The level of detail ranged from general to specific, with all states providing some sort of timeline for the assessment process. This theme was also evidenced by references to changes in policies and information. For example, states using external contractors to score their alternate assessments provided information about the scoring contractor and handling the process. Also, policies on such areas as security of results or violations of the code of conduct were included for two states. West Virginia noted, “Any breach of security, loss of material or other deviation from acceptable security procedures shall be reported immediately...” Virginia and North Carolina both addressed scoring appeals.

Surprisingly less evident was the amount of attention given to training and assistance to those involved with alternate assessment. This may be due to states’ focus on training during the first couple of years of implementation. There is evidence, however, that states continue to recognize the need for both training and assistance with materials. Documents also reflected the inclusion of a variety of constituents in the

training and a range of providers. For example, it was noted in a Virginia document, “Parent training will be conducted through Parent Resource Centers in school divisions.” Other recipients included special education teachers, principals, test coordinators, scorers and “other interested individuals.” Training was noted as being provided by principals, test coordinators, the state Office or Department of Special Education, and local school boards on such topics as proper administration, scoring, content standards, and procedures.

Accountability. Accountability has been described as a systematic means of collecting data, analyzing it, and using it to assure those inside and outside the educational system that schools are advancing in desired directions (CED, 2001; Elliot et al., 2001; Elliott & Thurlow, 2000; Hill & DePascale, 2003; Thurlow, Elliott et al., 1998). While most states focus on system accountability, some address student accountability as well. Researchers agree, however, that in order to produce meaningful outcomes for students, assessment results must be included along with those of their nondisabled peers (Burgess & Kennedy, 1998; Thompson & Thurlow, 2003; Thurlow et al., 1996).

This theme contained two subcategories: results and decisions. Data reflecting this category and the subcategory “results,” in particular, was found for all states on Principle 1 (All Students) or Principle 5 (Subgroup Accountability) of their NCLB Consolidated State Application Accountability Workbook. As such, many of the responses were very similar, noting that “all results” from the general assessment and alternate assessment “are included as part of the AYP equation.” A unique response was noted in the Virginia workbook related to this category:

Effective with the 2002-2003 academic year, the USED has directed Virginia to limit to 1% the number of scores from these alternate assessments for children with the most severe cognitive disabilities that can be counted as proficient in AYP calculation. As a statement of public record, let it be clear that Virginia is “agreeing” to this directive under protest and only because the USED has made it clear it is mandating it.

A notable variation included one reference not tied to NCLB or AYP, which addressed the intentions to include alternate assessment results in the accreditation of schools.

Improvement. Alternate assessment has been in place only a short time and the number of students who participate may be as small as one half to two percent (Thurlow et al., 1996; Warlick & Olsen, 1999). This may explain the paucity of literature regarding its effect on student, program, and system outcomes. Research has begun to assess these issues, particularly student outcomes and the alignment of alternate assessments with curriculum and instruction (Browder, Flowers et al., 2002; Browder et al., 2003), but state educational leaders have shifted their focus to the procedural mandates of NCLB with emphasis on system outcomes.

However, several exceptions to this emphasis were identified in the data. For example, one state noted that it would review its procedures every five years, or as necessary, to ensure that systems addressed the needs of all students. Particular attention was given in several states to the importance of focusing on a critique of statistical methodology to ensure that “decisions resulting from these procedures are reliable and valid....” Or, that “assessments must yield reliable and valid information that leads

directly to student learning and improved instruction.” Additional data dealt with monitoring of outcomes.

Inclusion. Of the 475 themes noted in state policy, 74 related to inclusion.

Analysis of the text revealed that the focus of state policy in this area is on (a) inclusion in statewide assessments, (b) exceptions to inclusion, (c) exemptions from inclusion, (d) prior exclusion of students from statewide assessments, and (e) inclusion across settings. Statements such as “This means there can be no exemptions from State Mandated Assessments” and “The only exceptions are those few students who are able to complete neither the regular, nor the alternate assessments, even with allowable accommodations” are examples from this category. Also noted was, “... with few exceptions, all students in Kentucky must participate in the regular assessment or the alternate portfolio.”

Several states spoke of inclusion in their statements regarding the purpose of the alternate assessment. For example, several Virginia documents included the following statement: “The purpose of the VAAP is to evaluate the performance of students who have traditionally been exempted from state assessment programs.” All of the states included guidelines for the inclusion of students receiving homebound or hospital instruction: “A student eligible for participation in the Alternate Portfolio Assessment Program who is receiving instruction in home/hospital settings shall participate in the Alternate Portfolio unless the student has an injury or illness verified by a physician....” The policies also addressed other settings in terms of which type of assessment students participated in - the general assessment or alternate assessment. The following settings were identified:

- Special education schools
- Home/hospital setting

- Residential settings (excluded)
- Non-graded programs
- Alternative schools
- State-operated programs
- Department of Corrections
- All public schools

Five states made reference to guiding principles for the inclusion of students in statewide assessments, including the alternate assessment. The notion that “all children can learn” and “all students are included” were most evident.

Access. This theme was evident across the states in language related to access to (a) the general curriculum, (b) a functional curriculum, and (c) instruction. Statements dealing with access to the general curriculum were generally tied to the assessment’s relationship to state content standards or the state’s standard course of study. One state’s documents provided evidence that its assessment was “designed for students who are pursuing a functional curriculum” However, the majority included language that tied the assessment to the general curriculum. For example, “Participation in assessments should also promote access to the general curriculum ...” and “[student’s program] ...based on the general education curriculum and such life domain areas as vocation, recreation-leisure, and personal management.”

References to instruction appeared in terms of the range or nature, location, or focus of instruction. This is evidenced in the following properties and dimensions of the theme:

- systematic instruction
- ongoing daily instructional programs
- practical instructional activities
- individualized instruction
- appropriate instruction
- community based instruction
- instruction of targeted skills

- embedded instruction
- extensive instruction
- instruction in real environments
- authentic instruction or real-life context

Although “instruction” was a subcategory of the major theme “access,” the frequency of its inclusion in policy documents seems to suggest that states are committed to ensuring meaningful learning experiences for all students.

Reporting. Reporting was the only theme that persisted through all levels of the coding process. Although some of the other major themes (Accountability and Improvement) were among the first codes identified, at various times during the constant comparison, resorting, and analysis, they were subsumed by other categories such as Responsibility and Impact. The theme of Reporting centered specifically on reporting alternate assessment results. Thus, data focused on results related to participation (number and percent) and performance (level, proficiency, content area). Further, much of the language dealt with (a) who received the reports (parents, students, USED, state DOE, the public, school personnel); (b) who provided the reports (DOE); (c) the report content (proficiency results, assessment results, student performance, subgroup performance, participation, scores); and (d) the source of the information (report cards, performance reports, annual reports). The most frequently noted statement in the documents on reporting, however, was taken from Section 300.138 of IDEA 97

Report to the public, with the same frequency and in the same detail as reports on assessments are issued for children without disabilities participating in regular and alternate assessments, and performance results on regular and alternate assessments, including both aggregated and disaggregated data.

Summary of Themes

This last stage of the study attempted to discover what additional themes emerged from the analysis of state legislative, policy, and procedure documents that drive alternate assessment. A grounded theory process of open, axial, and selective coding was used to accomplish this. Initial coding revealed evidence of several of the core principles espoused in the literature. Namely, the major categories of Reporting, Accountability, and Improvement were those also identified by Quenemoen, Thompson et al. (2001) and Thurlow et al. (2001) as indicators of best practice. In this study, however, Access and Inclusion emerged as major categories, while they are both identified as characteristics of the core principle “All students” in the literature (Quenemoen, Thompson et al., 2001; Thurlow et al., 2001). Not surprising was the emergence of Resources as a dominant category, as constituents are still struggling with implementing the mandates of NCLB and refining alternate assessment programs, mandated by IDEA. This is represented visually in Figure 1. Themes common to the literature and policy documents appear in bold.

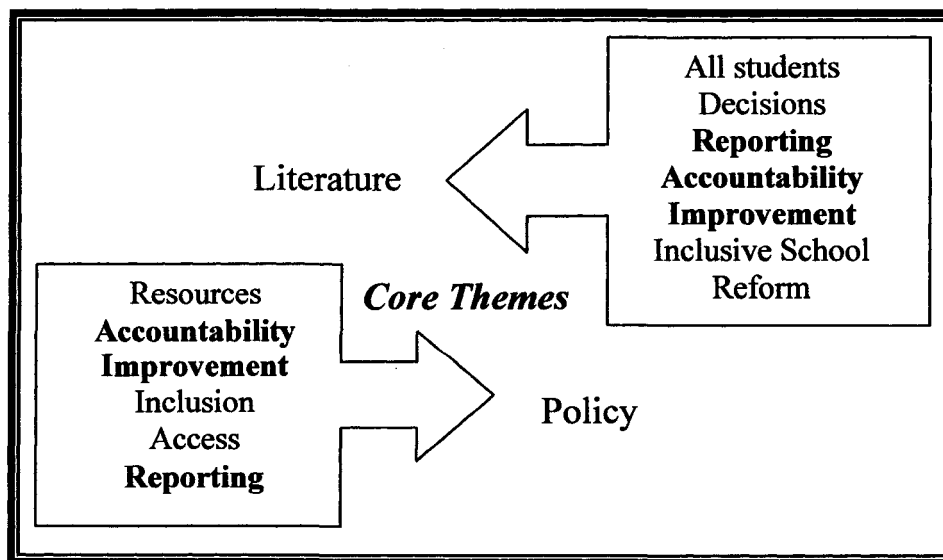


Figure 1. Core themes of inclusive assessment and accountability in the literature and state policy documents.

The final stage of grounded theory calls for selective coding to build “a ‘story’ that connects the categories” (Creswell, 1998, p. 150). Using diagrams, early memos and notes generated during the ongoing comparisons, relationships between the categories were developed to yield what seems to be a plausible central phenomenon - *The impact of policy on students with the most significant disabilities*. The data offer that, in spite of states’ provision of resources in response to legislative mandates accountability for student, program, and system improvement has only been minimally addressed. What has been observed is an increase in the inclusion of students with the most serious disabilities in statewide assessments and growing attention to access to curriculum and instruction, as reported in state documents.

Summary of Analysis

A priori categories based on existing research served as a starting point for this analysis. By developing a set of indicators that were tested to confirm their representativeness of these categories, evidence of core themes in states’ policy documents was identified. To varying degrees, all the themes were found in every state’s documents. More than half of the themes were noted in policy and procedure documents and manuals. Across all states, the most prevalent themes were “All Students” and “Decisions,” whereas the least prevalent were “Accountability” and “Monitoring.” This finding is not inconsistent with the sequence states followed in fulfilling the mandates of IDEA to include students with disabilities in statewide assessments. Initially, states developed alternate assessments that would be suitable for students with significant

disabilities who were previously excluded from general assessments. Then, participation guidelines were developed for IEP teams to use during their decision-making process. Systems for scoring and reporting alternate assessment results were in place for most states, however, reliability and validity measures had yet to be completed. As such, including these results in state accountability indices had not been attempted in most states. Further, states were just beginning to assess the impact of the assessment on student learning when the mandates of NCLB were enacted; bringing such evaluations to a halt.

Policy documents across states showed more differences than similarities though there were some like qualities. The main similarity was that a large number of coded evidence was not indicative of widespread presence of the themes but that the indicator was frequently repeated. Because of variations in the number and types of documents, percentages were used to analyze the existence of the themes in state documents. Except for West Virginia and Kentucky, there was a large spread between the percentage of indicators observed most and those observed least. The most noticeable difference was where states placed emphasis within the documents.

At the outset it was believed that additional themes, beyond those espoused in the literature, could be found in these same documents. Using a grounded theory approach of open and axial coding, such was found to be the case. Although the themes "Reporting," "Accountability," and "Improvement" came out, consistent with the literature, "Resources," "Inclusion," and "Access" also emerged. Through a process of selective coding, developing relationships between the themes, a plausible explanation of these findings was formulated. That is, in spite of states' provision of resources in response to

legislative mandates, accountability for student, program, and system improvement has only been minimally addressed. What has been observed is an increase in the inclusion of students with the most serious disabilities in statewide assessments and growing attention to access to curriculum and instruction, as reported in state documents. This central phenomenon will be further discussed in the next chapter on findings, recommendations, and implications.

CHAPTER V - Findings, Recommendations, and Implications

The purpose of this study was to examine alternate assessment policy from six states with an emphasis on the existence of the core principles of inclusive assessment and accountability systems reported in the literature. The study employed a mixed design using content analysis and grounded theory to make inferences and identify emerging themes in state alternate assessment policy. After completing steps to address the reliability of the codebook developed in Phase One of the research, legislation, policy, and procedure documents were collected for Kentucky, Maryland, North Carolina, Tennessee, Virginia, and West Virginia. The Second Phase of the study involved coding these documents to determine the presence of the core principles (in quotes) of inclusive assessment and accountability systems espoused in the literature. Lastly, a grounded theory process was used to identify additional themes (in italics) that emerged from the data.

This chapter presents a summary of the findings as well as recommendations and implications for future research and policy. Then, closing statements are offered.

Findings

Evidence of Core Principles

Evidence of all of the core principles was found in state policy documents. Across the states, the most prevalent theme was “Decisions” closely followed by “All Students,” with 26% and 25% of the coded segments expressive of these themes, respectively. The

next highest coded theme was “Reporting” (20%), followed by “Inclusive Reform” (14%), “Improvement” (8%), and “Accountability” (6%). The themes were ranked in varying order across the individual states; however, four of the six states ranked “Decisions” and “All Students” highest. Notably, for five of the six states “Improvement” and “Accountability” were ranked the lowest.

If one were to classify the themes as process versus product, the former classification might contain “Decisions,” “Reporting,” and “Accountability,” while the latter “All Students,” “Inclusive Reform,” and “Improvement.” Considering this in light of the contention in the literature that these principles must exist to support inclusive assessment and accountability systems, there is a balance of sorts between the two classifications. The study revealed a similar balance in the presence of the themes in policy documents. The themes that emerged in this study consistent with process were *Accountability, Resources, and Reporting*. Product, on the other hand, was represented by *Improvement, Inclusion, and Access*. As such, 219 (52%) of the coded segments could be classified as process, while 203 (48%) as product. On the surface, this would suggest that states’ policy is aligned with best practice. However, many of the references to the principles in the documents coded appeared within three contexts: What states (a) intend to do, (b) are in the process of doing, or (c) will do contingent on legislative revisions.

Similarities and Differences Across States

Similarities and differences seemed to be influenced by the nature of the assessment program (i.e., a single accountability system, implementation date of alternate assessment, degree of control of legislative bodies). Documentation from Kentucky, for example, was generally in the form of single documents that addressed the entire

assessment system. The language in the state's documents consistently reflected the inclusion of the Alternate Portfolio Program results in all accountability measures since its inception. The remaining states offered a majority of separate documents for each component of their statewide assessment system each addressing the areas of implementation, participation, scoring, and reporting.

Similarities in the use of very general references were noted in legislative documents across all states except for North Carolina. Together, its Board of Education Manual and general statutes contained the greatest percentage of references reflective of the core principles for documents categorized as legislation. One might speculate that behind these documents lies the real source of power over educational practices in the state. There was no evidence in the documents of such pronounced involvement for any of the other states.

Another similarity was noted in the areas of scoring and reporting. While states tended to separate procedural guidance on scoring and reporting for alternate assessment when addressing accountability, "all assessments" was used in all states' policy except for West Virginia. In almost all instances when the state's general assessment, the WESTEST, was mentioned, so was the alternate assessment. This held true for language in the State of West Virginia Consolidated State Application Accountability Workbook in sections addressing reporting and accountability.

The greatest number of core principles was identified in Virginia documents. This seemed due to the repetitive nature of the documents. Thus, information in the policy and procedure manuals was repeated numerous times within and across documents in an apparent attempt to prevent misinterpretation. At the same time, one could argue that the

complexity of the assessment calls for such repetition.

Emerging Themes

It was anticipated at the outset that the themes identified in the literature using pre-existing categories were not the only ones present in state documents. To explore this assumption, a second level of qualitative analysis was used to answer Research Question 4, “What additional themes emerge from analysis of the state legislative, policy, and procedure documents that drive alternate assessment programs?” Grounded theory, described by Strauss and Corbin (1998) as “theory that was derived from data, systematically gathered and analyzed through the research process” (p. 12), was used to examine and develop an understanding of state policy governing alternate assessment.

The process involved three phases, which were not linear but involved a constant comparison of the raw data. Through an initial process of open coding, the data were coded, sorted, and analyzed to a point of saturation where no new information was obtained. This process yielded a set of six categories for which subcategories, properties, and dimensions were also identified. Once these categories were identified, evidence of their relationship to one another was sought. This process of axial coding, or relating the categories to their subcategories, served as the basis for the final step - selective coding.

After identifying this set of categories and examining the relationships among them, a central phenomenon was derived. Once this was done, statements were developed to explain the relationship of the major categories to this central concept. To arrive at a central concept, diagrams were created based on early memos using a process of sorting and rearranging. What resulted as the central phenomenon was, *the impact of policy on students with the most significant disabilities*. Based on the emerging themes, *Resources*,

Accountability, Improvement, Inclusion, Access, and Reporting, the following theoretical explanation is offered:

In spite of states' provision of resources in response to legislative mandates, accountability for student, program, and system improvement has only been minimally addressed. What has been observed is an increase in the inclusion of students with the most serious disabilities in statewide assessments and growing attention to access to curriculum and instruction, as reported in state documents.

This explanation was the result of numerous attempts to diagram the results in a way that was logical, consistent, and explanatory; not forced.

Recommendations and Implications for Future Research

Fit of the Theory

Although not expressly stated, it was clear that the strongest condition influencing states' inclusion of students with the most significant disabilities in their assessment and accountability systems is the requirement to comply with federal mandates.

Unfortunately, compliance does not ensure results. Exploration of this issue offers several opportunities for future research.

One such study could seek to determine how well the theoretical explanation offered in this study fits the evidence. Strauss and Corbin (1998) recommended "validating" the results of grounded theory. By this they mean, "The theory emerged from data, but by the time of integration, it represents an abstract rendition of that raw data. Therefore, it is important to determine how well that abstraction fits with the raw data and also to determine whether anything salient was omitted ..." (p. 159). The authors suggested several ways to accomplish this. Among them, going back and

comparing the resulting scheme to the raw data by doing a high-level comparative analysis. Alternatively, the researcher could share the “story” with other respondents or representatives and ask them how well they believe it fits their case. Obtaining feedback from individuals directly involved with crafting or overseeing state policy on assessment and accountability would serve not only to validate the results as Strauss and Corbin suggested, but also lend credibility to the interpretation and explanation of the findings (Denzin & Lincoln, 1998; Miles & Huberman, 1994). Nevertheless, although a follow-up study involving interviews with alternate assessment specialists could provide this sort of validation, the current shifting environment has political implications that may affect individuals’ comfort with or ability to discuss state assessment and accountability policy. Thus, the initially proposed design for this research called for such interviews to substantiate the findings of the content analysis. However, an alternate methodology was chosen due to challenges in obtaining respondents.

Similarity of the Findings

A subsequent study could include the same methodology using different states. Although a great deal of variation was found in the documents across states, the delimitations imposed by the study’s design and the nature of qualitative research affect to some extent the ability to make generalizations to other states. In spite of this limitation, it is expected that the emphasis on compliance extends to other regions in spite of the early evidence that states are beginning to move beyond process toward an emphasis on student learning.

Although it was not considered at the time the six states were selected for inclusion, it was later noted that diversity in alternate assessment characteristics made for

a more representative sample. The *2003 State Special Education Outcomes: Marching On* (Thompson & Thurlow, 2003) provided information on states reflecting the use of multiple alternate assessments, the variety of alternate assessment scorers, and the range of standard-setting techniques states use, as well as emerging issues such as computer-based assessment. Tennessee and North Carolina were among the states offering two types of alternate assessment. Sixty-two percent of states used the same descriptors for their general and alternate assessment, according to Thompson and Thurlow. West Virginia was among the 32% that used different achievement-level descriptors. The authors further reported that most alternate assessments were scored by teachers (from other districts or the students' teachers), followed by test contractors and state educational agencies. All of these scoring approaches were found in the sample states. Thompson and Thurlow also reported current and emerging issues, among them out-of-level testing options. North Carolina and West Virginia reported including their out-of-level results in the calculation of assessment participation rates (Edwards, 2004).

Change in State Policy Over Time

The areas of assessment and accountability are moving targets and this study provided a glimpse of only one moment in time. Since initiating the study in the winter of 2003, changes in states' policies have occurred. Already Kentucky, Tennessee, Virginia, and West Virginia as well as other states, have submitted requests to amend their accountability plans. Maryland and West Virginia have implemented their revised alternate assessment. Maryland plans to set its passing standards for the Alt-MSA in Grades 4, 6, and 7 during summer 2004. The West Virginia Alternate Assessment has gone from using 11 Instructional Goals and Objectives as the bases for evaluating student

performance to the use of 8 mandated Content Standards and Objectives. Thus, it might be useful to repeat this study using the same states after they have had an opportunity to collect data based on revised assessment criteria. This could help determine whether such revisions are based solely on federal mandates or if they are driven by assessment results and classroom practices.

Recommendations and Implications for Policy

Consider the Message

If states are able to do what they have been mandated to do, their policy documents suggest that they should be able to achieve the type of a balance between process and product that aligns with best practice and yields desirable consequences for students with the most serious disabilities. Unfortunately, the message in current state policy weighs heavily in favor of process. This appears consistent with an increased sense of urgency to go beyond simply including and reporting the results of students with disabilities in statewide assessments. Although the number of students participating in alternate assessment programs may be small, only ranging from .5%-1% in most cases, educational leaders face the challenge of using performance results for this group of students in calculating the adequate yearly progress called for under NCLB.

Understandably, states are still attempting to ensure that the mechanisms they use to do so are reliable and yield valid results, as the stakes have increased.

Continue to Provide Supports

As Hord (cited in Meister, 2000) explained, change “must be viewed as a journey by individuals who have highly personal views and levels of understanding, which are evident in the different ways that they develop through a change endeavor” (p. 28).

Although the focus of this study was state policy, the resulting central phenomenon, the impact of policy on students with the most significant disabilities, links directly to the classroom. If the goal of NCLB is to not only include students with the most significant disabilities in state assessment and accountability systems but also to ensure that these students receive a quality education, policymakers must ensure shared accountability among educators, related service providers, students, and their families. These are the individuals for whom the outcomes are most personal. As such, they should not be the ones impacted most by negative consequences. What we should see over time are changes in practice and instruction that result in student learning, not concentrated efforts to meet AYP targets that may be unrealistic for students with significant disabilities.

States had begun to consider the technical adequacy of their alternate assessment measurement and scoring processes prior to the implementation of NCLB. This review seems to have taken a back seat to the requirement to include alternate assessment scores in their AYP calculations. Unlike norm-referenced measurements, which are used to discriminate among students at different achievement levels, criterion-referenced measurements are intended to determine whether students have achieved the established standard (Gall et. al., 1996). Further, the basis for establishing reliability for the two types of referenced measurements is different. Policymakers must now question the feasibility and monitor the impact of using such criterion-referenced results in AYP determinations.

Monitor the Impact of Legislation and Results

The findings of this study suggest that the states have to work quickly through this process of compliance or at least simultaneously increase their attention to the product side of the scale. Only by doing so might they realize a system of assessment and

accountability consistent with research on best practice. By tipping the balance in favor of positive consequences (*Improvement, Inclusion, Access*), states should see the outcomes they desire. The challenge will be to do so in this shifting policy environment. For example, Virginia passed a resolution challenging the soundness of the NCLB mandates and, like North Carolina, has decided to use two separate ratings for schools (U.S. Department of Education, 2004). We are seeing schools deemed as failures under NCLB while simultaneously exceeding state standards for accreditation or considered passing (Hoff, 2004). Is the goal of this legislation to prescribe the amount of improvement students with the most significant disabilities should achieve or to ensure that they receive a high quality education?

On the federal side, we have seen some relaxation in the standards related to alternate achievement standards as well as flexibility in calculating participation rates. Since December 2003, the U.S. Department of Education has made four changes to the No Child Left Behind Act mandates. On December 9, 2003, the Department finalized a regulation that allows states to use alternate assessments based on alternate achievement standards for students with the most significant disabilities, which will enable them to meet AYP requirements. At the same time, guidance has been provided that established a 1.0% cap on the number of proficient and advanced scores that could count toward AYP (Improving the Academic Achievement, 2003). On February 29, 2004, U.S. Secretary of Education, Rod Paige, announced two policies to ensure flexibility and accountability under NCLB for students with limited English proficiency. Additionally, on March 29, 2004 new policies were announced for calculating rates of students who participate in an assessment program. All of these occurrences have implications for state policymakers

and administrators as they continue to operationalize the current shifting policy.

Closing Statements

It may be too soon to say what is best practice for assessing students with the most significant disabilities; however, it is not too soon to know that states must be held accountable for these students' achievement. When this group of students is excluded from statewide assessments, they are also excluded in policy decisions that affect them. Alternate assessment has only been in place for most states since July 2000. Just as states were beginning to move beyond procedural issues of implementing this assessment the No Child Left Behind Act of 2001 was enacted, presenting a host of new challenges. However, because of the potential impact on general education, all constituents have an obligation to ensure a meaningful education and positive outcomes for all students.

Evidence from this research that states are providing access to the general curriculum and including students with the most significant disabilities in assessment systems is encouraging. Now, the results of this activity must be carried out in accountability data called for by NCLB. Ongoing as well as recent correspondence to Chief State School Officers (R. Simon, personal communication, March 2, 2004) serves as a reminder that the President and the U.S. Department of Education do not intend to waive the issues of adequate yearly progress as a measure of improvement. As states address AYP requirements, they must reassess the alignment between their general and alternate assessment systems. In this study, this relationship was described in state policy as linkage, alignment, and extensions. NCLB requires states to either align alternate assessment with state content standards or develop alternate academic achievement standards (also referred to as alternate achievement standards) that differ in complexity.

There is a 1.0% cap, however, on the number of scores based on these alternate achievement standards that may be used in AYP calculations.

State education documents reflect leaders' belief that all students must be included in assessment and accountability systems for students to achieve positive outcomes. Yet, no matter how successful states have been in doing so, they are feeling the impact of current federal mandates. Even states such as Kentucky and Maryland that have been assessing and including students with severe disabilities in statewide assessment and accountability systems for over 10 years have revisited their programs to meet the most recent and still unfolding aspects of NCLB. In sum, despite the challenges imposed by this, and preceding legislation, results of this study indicate that students with the most significant disabilities are benefiting from inclusion in state assessment and accountability systems.

APPENDICES

Appendix A

Guidelines for Coding

All students included

1.
 - a. **Label:** AL-STD
 - b. **Characteristic:** Alternate assessment is aligned with state standards held for all students through some process of extension, expansion, access, or other high performance bridge to state content standards.
 - c. **Indicators:** General assessment [name], state standards, extension, access, linkage, congruent with.
 - d. **Criteria:** Statement that explains the relationship between what the alternate and general assessments measure; **OR** statement that addresses the relationship between the alternate and general assessment.
 - e. **Examples:** Document states that ...
 - i. All students with disabilities are included in the assessment system in some way.
 - ii. ... performance in the alternate assessment program is tied to state content standards.
 - iii. ... no exemptions; no exclusions.

2.
 - a. **Label:** AL-OPT
 - b. **Characteristic:** Alternate assessment options are available across all components of state or district assessment system.
 - c. **Indicators:** State-operated programs, juvenile justice system.
 - d. **Criteria:** Evidence that alternate assessment is available as an option for all students who receive educational services in any setting (i.e., state-operated programs, juvenile justice system, home-bound, hospital).
 - e. **Examples:**
 - i. ... students are eligible for the alternate assessment across all settings.
 - ii. ... students receiving educational instruction in any setting may be eligible to participate in the alternate assessment.
 - iii. ... regardless of school placement, a student may be eligible to participate in the alternate assessment.

3.
 - a. **Label:** AL-STR
 - b. **Characteristic:** Alternate assessment options promote the use of a variety of valid authentic performance-based assessment strategies aligned to standards, allowing all students to show what they know and are able to do.
 - c. **Indicator:** Portfolios, checklists, observation, test, body of evidence, performance assessment, traditional test, pen/paper test, computer test.

- d. **Criteria:** Evidence that the assessment uses such instruments as portfolios, checklists, or observation to determine each student's proficiency in skills or tasks based on content standards.
- e. **Examples:**
 - i. Student proficiency in the alternate assessment program is determined using observation and checklists.
 - ii. A body or collection of evidence is collected to assess student performance in the alternate assessment program.

Decisions

1.

- a. **Label:** DE-PER
- b. **Characteristic:** State policies include an estimate of percent of students expected to participate in alternate assessments, as general guidance to help teams understand the need to include majority of students in the general assessment.
- c. **Indicators:** Estimate percentage participating, who participates, participation in the assessment.
- d. **Criteria:** Statement exists regarding number or percentage of students expected to participate in the alternate assessment program.
- e. **Examples:**
 - i. Approximately 1% of the students eligible for assessment will take the alternate assessment.
 - ii. Less than 2% of the total population is expected to take the alternate assessment.
 - iii. **NOT** – statement regarding federal requirements for participation in assessment program(s) or reference to federal policy estimate.

2.

- a. **Label:** DE-ABL
- b. **Characteristic:** Decisions based on ability of student to take the general assessment with or without accommodations.
- c. **Indicators:** IEP decision, IEP team, general assessment, student ability.
- d. **Criteria:** Evidence that the general assessment is the starting point in the decision process related to participation.
- e. **Examples:**
 - i. Decisions are not based on placement (i.e., self-contained, resource room, etc.).
 - ii. Decisions are not based on student working toward standards in the general curriculum.
 - iii. Decisions are not based on what the student is expected to do or how they are expected to perform.

3.
 - a. **Label:** DE-TML
 - b. **Characteristic:** There is a clear explanation of the process for compiling data on individual alternate assessment decisions to be used in planning and ordering materials, and for training purposes.
 - c. **Indicators:** Date(s), training, ordering material, ordering equipment/ supplies, submission due date, planning schedule, calendar, administration dates.
 - d. **Criteria:** Evidence that guidelines are provided to individuals involved in the alternate assessment regarding timelines for administration and completion as well as dates/schedule for such things as training and ordering materials/ equipment.
 - e. **Examples:**
 - i. The alternate assessment will be conducted annually for all eligible students in the 3rd, 5th, and 8th grades ...
 - ii. Portfolios must be submitted no later than the third week of March.
 - iii. Training will be conducted ...

Reporting

1.
 - a. **Label:** RE-TNP
 - b. **Characteristics:** State and district reports include the number and percent of students participating in the alternate assessment and each subtest within it.
 - c. **Indicators:** Number of students reported, percent of students reported.
 - d. **Criteria:** Statement exists that the total number and percent of students participating in the alternate assessment is publicly reported as well as the number and percentage of those participating in each content area (Math, Science, Social Studies, English).
 - e. **Examples:** Evidence of a summary or score report of some type that reveals ...
 - i. A total of 3029 students participated in the alternate assessment.
 - ii. .5% of all students assessed participated in the alternate assessment.
 - iii. Math ($n=239$), science ($n=241$), etc.
2.
 - a. **Label:** RE-SCO
 - b. **Characteristics:** State and district reports include scores for students participating in the alternate assessment, although scores may be in disaggregated form.
 - c. **Indicators:** Scores included, report cards, reports, score reports, same frequency, same format.
 - d. **Criteria:** Reports include all scores for students participating in the alternate assessment.

- e. Examples:
 - i. Score reports include the scores of students with significant disabilities not able to participate in the general assessment even with accommodations.
 - ii. All assessment scores, including those for students with significant disabilities, are included.
 - iii. Alternate assessment scores are aggregated with those of students taking the general assessment [name].
- 3.
- a. Label: RE-VAL
 - b. Characteristics: Rubrics are developed to reflect research and best practice understanding of desired outcomes for students with significant disabilities. Validity measures include the study of face validity of the rubrics.
 - c. Indicators: Rubrics, checklists, descriptive continuum, validity, reliability, face validity.
 - d. Criteria: Documentation exists regarding the purpose of the assessment and how validity was established.
 - e. Examples:
 - i. Each portfolio is scored using a rubric.
 - ii. Rubrics were developed to score each portfolio.
 - iii. Validity of the scoring process was assessed using
- 4.
- a. Label: RE-APP
 - b. Characteristics: Scoring and reporting processes include a detailed approach for administration, with clearly defined performance standards, scoring and recording procedures, and reliability checks built into the process.
 - c. Indicators: Reporting guidelines/procedure, methodology, guidance, scoring process, trained scorers, standardization, standard setting.
 - d. Criteria: Documentation exists regarding how the assessment will be scored, who will score it, how the information will be compiled, and how the information will be reported.
 - e. Examples:
 - i. Key stakeholders determined cut scores were using a standard setting process.
 - ii. Two individuals score each portfolio; a third may be required when scores are not adjacent.
 - iii. Each collection of evidence is scored independently by two scorers.
- 5.
- a. Label: RE-CON
 - b. Characteristics: The confidentiality of individual participants is ensured by not publicly reporting any disaggregated data that may compromise

- student privacy.
- c. **Indicator:** Confidentiality.
 - d. **Criteria:** Reports show no evidence of disaggregated data that could be linked to a specific student.
 - e. **Examples:**
 - i. Student scores are not reported individually for schools having fewer than 10 students participating in the alternate assessment.
 - ii. ... ensure confidentiality and the rights of the individual.

Accountability

- 1.
 - a. **Label:** AC-INT
 - b. **Characteristic:** Alternate assessment scores are integrated into the adequate yearly progress accountability index as an equal factor with all other scores.
 - c. **Indicators:** Equal weight, count the same, totaled with.
 - d. **Criteria:** Documentation exists reflecting that performance data for all students (regardless of how they participate) are included in AND have the same impact as all other student performance data in the accountability index.
 - e. **Examples:**
 - i. All scores are figured in (or calculated as part of) the state's accountability index.
 - ii. (negative example) All scores are counted but alternate assessment scores are reported separately.

Improvement

- 1.
 - a. **Label:** IM-MON
 - b. **Characteristic:** There are monitoring and evaluation processes that include gathering of data on use of alternate assessments to improve practices at the student, school, district, and state levels, aligned to emerging research and best practices.
 - c. **Indicators:** Improve practice, monitoring, evaluation, revisited, revised.
 - d. **Criteria:** Evidence of a process for monitoring and evaluating alternate assessment data that indicates how the results of such activities will be used.
 - e. **Examples:**
 - i. Alternate assessment results are reviewed, and appropriate personnel in the district notified of any alerts.
 - ii. Alternate assessment data are evaluated for improvement across districts and the state.
 - iii. Alerts are used to provide feedback and strategies for improvement.

2.
 - a. **Label:** IM-TRA
 - b. **Characteristic:** Based on monitoring and evaluation data, and emerging research and best practice indicators, training on alternate assessment is in place for IEP teams and other key personnel including principals, counselors, school psychologists, and others.
 - c. **Indicator:** Training, train-the-trainer, inservice, education, question and answer.
 - d. **Criteria:** Monitoring and evaluation data exists that is incorporated into training for IEP teams and any others involved in the alternate assessment program.
 - e. **Examples:**
 - i. Alternate assessment results are shared with school personnel and strategies discussed ...
 - ii. Administrators and key personnel will attend training sessions on alternate assessment updates.

Inclusive School Reform

1.
 - a. **Label:** IN-INS
 - b. **Characteristic:** Alternate assessment practices are embedded in standards-based instructional activities throughout the course of the assessment period across multiple settings.
 - c. **Indicators:** Indicators embedded, instructional activities, assessment period.
 - d. **Criteria:** Evidence exists that indicators, assessed via alternate assessment are those included as part of the student's instructional plan and apply across multiple settings with nondisabled peers.
 - e. **Examples:**
 - i. The portfolio shows evidence of the student's work throughout the school year.
 - ii. The collection of evidence is consistent with the student's instructional activities.
 - iii. IEP goals incorporate activities across a variety of setting, including those with nondisabled peers.
2.
 - a. **Label:** IN-XSE
 - b. **Characteristic:** Alternate assessment practices are integrated with other related efforts like inclusion models, transition planning, and best practice IEP development.
 - c. **Indicators:** Inclusion, transition, IEP planning, graduation, diploma.
 - d. **Criteria:** Evidence that alternate assessment policies, guidelines, and training are addressed or included as part of efforts across all educational settings and levels.

e. Examples:

- i. Transition or graduation plans are considered in the development of IEP goals.
- ii. Students participating in the alternate assessment are those not eligible for a standard diploma.

Appendix B

Coding Instructions

(To be used in conjunction with Coding Form shown in Appendix C)

Step 1: Read the entire document.

Step 2: Record bibliographic and informational data in the space provided on the coding form. Use one form per document.

Step 3: Analyze the document to identify the presence of the 16 categories specified.

To do so, review each line of the document. For each document, highlight or enclose the phrase, sentence, paragraph, or page(s) that demonstrates evidence of a specific indicator in brackets. Label the bracketed text with the code that identifies the theme evidenced. For example:

[The IEP team may determine that the disability is serious enough that the student cannot participate in standard statewide test administrations with or without accommodations. These students must be administered the state-designated alternate assessment. The basis for the decision must be documented using current longitudinal data and must not be the result of social, cultural, and/or economic differences.] DE-ABL

Step 4: Place a tally in the “Evidence” box if the indicator is present. Indicate the corresponding page number from the text. Place a tally in the “In Progress” box if evidence exists that the indicator has been considered or is being addressed. Also, indicate the page number. After analyzing the entire document, check the “No Evidence” box for those indicators for which no evidence exists. In cases where

evidence exists or is in progress, note on the coding form each key idea or phrase that corresponds to the indicator in the space provided below each item.

Appendix C
Content Analysis Coding Form

Document: _____			
Source: _____			
Publication/Retrieval Date: _____		Coder : _____	
Document Type: _____ Legislation _____ Policy _____ Report _____ Manual			
	Evidence	No Evidence	Comments (i.e., in progress)
1. Alternate assessment is aligned with state standards. AL-STD			
2. Alternate assessment options are available across all components of state or district assessment system. AL-OPT			
3. Alternate assessment options promote the use of a variety of performance-based assessment strategies. AL-STR			
4. State policies include an estimate of percent of students expected to participate in alternate assessments. DE-PER			
5. Decisions based on ability of student to take the general assessment with our without accommodations. DE-ABL			
6. There is a clear explanation of the process for compiling data on individual alternate assessment decisions. DE-TML			
7. State and district reports include the number and percent of students participating in the alternate assessment and each subtest within it. RE-TNP			
8. State and district reports include scores for students participating in the alternate assessment. RE-SCO			
9. Rubrics are developed and studied for face validity. RE-VAL			
10. Scoring and reporting processes include a detailed approach for administration, with reliability checks built into the process. RE-APP			

	Evidence	No Evidence	Comments
11. The confidentiality of individual participants is ensured. RE-CON			
12. Alternate assessment scores are integrated into the AYP accountability index as an equal factor with all other scores. AC-INT			
13. There are monitoring and evaluation processes to improve practices. IM-MON			
14. Training on alternate assessment is in place for IEP teams and other key personnel. IM-TRA			
15. Alternate assessment practices are embedded in standards-based instructional activities across multiple settings. IN-INS			
16. Alternate assessment practices are integrated with other related efforts. IN-XSE			

Appendix D

Documents Included in the Content Analysis

1. **Kentucky (11)**
 - a. **Kentucky Administrative Regulations**
 - b. **Kentucky Revised Statutes**
 - c. **Consolidated State Application**
 - d. **Kentucky's Consolidated State Application Accountability Workbook**
 - e. **Kentucky Board of Education Issue Briefs**
 - f. **Program of Studies for Kentucky Schools**
 - g. **2003 Commonwealth Accountability Testing (CATS) System Interpretive Guide**
 - h. **2002-2003 District Assessment Coordinator (DAC) Implementation Manual**
 - i. **Implementation Manual for the Program of Studies**
 - j. **The Kentucky Board of Education Annual Report**
 - k. **The Kentucky Board of Education Strategic Plan**

2. **Maryland (10)**
 - a. **Code of Maryland Regulations**
 - b. **Maryland Statutes**
 - c. **Consolidated State Application**
 - d. **Maryland State Department of Education Consolidated State Application Accountability Workbook**
 - e. **Implementation Procedures for Making Adequate Yearly Progress Determinations for No Child Left Behind**
 - f. **Requirements for Accommodating, Excusing, and Exempting Students in Maryland Assessment Programs**
 - g. **Alternate Maryland School Assessment (Alt-MSA) Test Administration and Coordination Manual**
 - h. **School Assessment (MSA): 2003 Posttest Specification Manual**
 - i. **Maryland State Improvement Grant Performance Report**
 - j. **Consolidated State Performance Report**

3. **North Carolina (13)**
 - a. **North Carolina Administrative Code**
 - b. **North Carolina General Statutes**
 - c. **North Carolina State Board of Education Policy Manual**
 - d. **North Carolina Department of Public Instruction Consolidated State Application**
 - e. **State Board of Education Consolidated Application Accountability Workbook**
 - f. **Testing Students with Disabilities: North Carolina Testing Program**
 - g. **Determining Composite Scores in the ABCs Model**

- h. Understanding Proficiency Statistics in Various Reports
- i. North Carolina Alternate Portfolio: Portfolio Development Designee/Teacher Handbook
- j. Student Accountability Standards Report 2002-2003
- k. State Board of Education Biennial Report 2000-2002
- l. North Carolina's Strategic Plan for Excellent Schools
- m. Testing Students with Disabilities Advisory Committee: Recommendations and Discussion

4. Tennessee (13)

- a. Tennessee Code
- b. State Board of Education Rules, Regulations and Minimal Standards for the Operation of the Public School System
- c. Consolidated State Application
- d. Tennessee Department of Education Consolidated State Application Accountability Workbook
- e. State Board of Education Professional Development Policy
- f. Tennessee State Board of Education Performance Model 2003
- g. Tennessee Alternate Portfolio Assessment Teacher's Manual
- h. Tennessee State Department of Education Special Education Manual - 2003
- i. Collecting Instructional Data for the Portfolio
- j. Suggested Scoring Procedures for the TCAP-Alt Portfolio Assessment
- k. Official Report of Findings and Recommendations of the Advisory Committee on Accountability and Testing to the Tennessee State Board of Education
- l. State Board of Education Master Plan for Tennessee Schools: Preparing for the 21st Century
- m. Tennessee Advisory Council for the Education of Students with Disabilities Annual Report

5. Virginia (11)

- a. Code of Virginia
- b. Regulations Governing Special Education Programs for Children with Disabilities in Virginia
- c. Virginia's Consolidated State Application
- d. Virginia Board of Education Consolidated State Application Amended Accountability Workbook
- e. Standards of Quality
- f. Standards of Accreditation
- g. Virginia Department of Education Procedures for Participation of Students with Disabilities in the Assessment Component of Virginia's Accountability System
- h. Virginia Alternate Assessment Program Implementation Manual
- i. Virginia Alternate Assessment Program Administrator's Manual
- j. Virginia Substitute Evaluation Program Procedural Manual

- k. **Virginia's State Improvement Plan for Special Education 1999-2004**
 - l. **2002 Annual Report on the Condition and Needs of Public Schools in Virginia**
 - m. **Board of Education Six-Year Plan: 2002-2003**
6. **West Virginia (9)**
- a. **West Virginia Code**
 - b. **Board of Education Policy Manual**
 - c. **Greenbook of Public Education Bills Enacted in Regular Session**
 - d. **Consolidated State Application**
 - e. **State of West Virginia Consolidated State Application Accountability Workbook**
 - f. **Students with Disabilities: Guidelines for Participation In the West Virginia Measures of Academic Progress (MAP)**
 - g. **West Virginia Alternate Assessment Skill Inventory**
 - h. **West Virginia Measures of Academic Progress; Alternate Assessment Administration Manual**
 - i. **West Virginia IDEA Improvement Plan**

Appendix E

Summary of Categories, Codes, and Indicators

Category	Code	Indicators
ALL STUDENTS	AL	
AL: Align – State Standards	AL-STD	General assessment [name], content standards, state standards, extension, access, linkage, linked to, congruent with
AL: Options	AL-OPT	State-operated programs, juvenile justice system, home-bound, hospital
AL: Variety of Strategies/ Instruments	AL-STR	Portfolios, checklists, observation, test, body of evidence, performance assessment, traditional test, pen/paper test, computer test
DECISIONS	DE	
DE: Percentage	DE-PER	Estimate percentage participating, who participates, participation in general assessment
DE: Participate –Ability	DE-ABL	IEP decision, IEP team, general, assessment, student ability
DE: Timeline	DE-TML	Date(s), training, ordering material, ordering equipment/supplies, submission due date, planning schedule, calendar, administration dates
REPORTING	RE	
RE: Total Number and Percent	RE-TNP	Number of students reported, percent of students reported
RE: All Scores	RE-SCO	Scores included, report cards, reports, score reports, same frequency, same format
RE: Valid	RE-VAL	Rubrics, check lists, descriptive continuum, validity, reliability, face validity
RE: Detailed Approach	RE-APP	Reporting guidelines/procedure, methodology, guidance, scoring process, trained scorers, standardization, standard setting
RE: Confidentiality	RE-CON	Confidentiality
ACCOUNTABILITY	AC	
AC: Scores Integrated	AC-INT	Equal weight, count the same, totaled with,
IMPROVEMENT	IM	
IM: Process improves practice	IM-MON	Improve practice, monitoring, evaluation, revisited, revised

Category	Code	Indicators
IM: Training	IM-TRA	Training, train-the-trainer, inservice, education, question and answer (Q&A)
INCLUSIVE SCHOOL REFORM	IN	
IN: Indicators - Instructional Plan	IN-INS	Indicators embedded, instructional activities, assessment period
IN: Efforts - Across Settings	IN-XSE	Inclusion, transition, IEP planning, graduation, diploma

Appendix F

Summary of Test Coding with Second Coder

Document (number of pages)	Based on coding form		Based on documents			Comment
	Agree (n=)	Disagree (n=)	Total coded indicators (TRC)	Total coded indicators (KC)	Number of matches	
1 (17p)						Disagreement re: indicators 1 & 3
Form	14	2				
Document			7	19	2	Disagreement re: indicators 2 & 4
2 (8p)						
Form	14	2				
Document			4	2	1	
3 (1p)						
Form	15	1				
Document			2	3	2	Disagreement re: indicators 6, 15, & 16
4 (9p)						
Form	13	3				
Document			15	15	7	
5 (4p)						
Form	16	0				
Document			4	5	3	
6 (11p)						
Form	15	1				
Document			5	6	4	
7 (2p)						
Form	15	1				
Document			0	1	0	Disagreement re: indicators 1, 6, & 7
8 (12p)						
Form	13	3				
Document			9	9	6	
9 (2p)						
Form	16	0				
Document			2	2	2	
10 (2p)						
Form	16	0				
Document			3	5	3	Problems especially with coding indicators 1, 2, & 6
Total	147	13	51	67	30	

References

- Ahearn, E. M. (2000). *Educational accountability: A synthesis of the literature and review of a balanced model of accountability* (Final Report). Alexandria, VA: Project FORUM, National Association of State Directors of Special Education.
- Almond, P., Quenemoen, R., Olsen, K., & Thurlow, M. (2000). *Gray areas of assessment systems* (Synthesis Report No. 32). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved October 23, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis32.html>
- Alper, S. K., Ryndak, D. L., & Schloss, C. N. (2001). *Alternate assessment of students with disabilities in inclusive settings*. Boston: Allyn & Bacon.
- Bechard, S. (2001). *Models for reporting the results of alternate assessments within state accountability systems* (Synthesis Report 39). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved October 23, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis39.html>
- Bolt, S., Krentz, J., & Thurlow, M. (2002). *Are we there yet? Accountability for the performance of students with disabilities* (Technical Report 33). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved February 5, 2003, from <http://education.umn.edu/NCEO/OnlinePubs/Technical33.htm>
- Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code*

development. Thousand Oaks, CA: Sage.

Browder, D., Ahlgrim-Delzell, L., Flowers, C., Karvonen, M., Spooner, F., & Algozzine, R. (2002, April). *How states define alternate assessments for students with disabilities*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Browder, D., Flowers, C., Ahlgrim-Delzell, L., Karvonen, M., Spooner, F., & Algozzine, R. (2002, April). *Curricular implications of alternate assessments*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

Browder, D. M., Spooner, F., Algozzine, R., Ahlgrim-Delzell, L., Flowers, C., & Karvonen, M. (2003). What we know and need to know about alternate assessment. *Exceptional Children*, 70(1), 45-61.

Burdette, P. J., & Olsen, K. (2000, July). *Alternate assessments: A medley of alternate assessments*. Lexington: University of Kentucky, Mid-South Regional Resource Center. Retrieved October 19, 2002, from http://www.ihdi.uky.edu/msrrc/Publications/alternate_alternates1.htm

Burgess, P., & Kennedy, S. (1998). *What gets tested, gets taught, who gets tested gets taught: Curriculum framework development process*. Lexington: University of Kentucky, Mid-South Regional Resource Center.

Carney, T. F. (1972). *Content analysis: A technique for systematic inference from communications*. Winnipeg, Canada: University of Manitoba Press.

Cohen, M., & Heumann, J. E. (2001, January 12). *Clarification of the role of the IEP team in selecting individual accommodations, modifications in administration,*

- and alternate assessments for state and district-wide assessments of student achievement* [Memorandum]. Washington, DC: U.S. Department of Education.
- Committee for Economic Development. (2001). *Measuring what matters: Using assessment and accountability to improve student learning*. New York: Author.
- Consortium on Inclusive Schooling Practices. (2001). *Determining policy support for inclusive schools*. Retrieved February 12, 2003, from <http://www.edc.org/urban/pathways.htm>
- Council for Exceptional Children. (2000). *Making assessment accommodations: A toolkit for educators*. Arlington, VA: Author.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- deFur, S. H. (2002). Education reform, high-stakes assessment, and students with disabilities. *Remedial and Special Education*, 2(4), 203-211. Retrieved February 4, 2003, from Academic Search Elite database via EBSCOhost.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (1998). *Strategies of qualitative inquiry*. Thousand Oaks, CA: Sage.
- Education Commission of the States. (n.d.). Retrieved November 7, 2003, from <http://www.ecs.org>
- Educational Policy Reform Research Institute. (n.d.). Retrieved September 10, 2003, from <http://www.eprri.org>
- Edwards, V. B. (Ed.). (2004, January 8). Quality counts 2004: Count me in [Special issue]. *Education Week*, 23(17).
- Elementary and Secondary Education Act of 1965, 20 U.S.C. 2701 (1965).

- Elliott, S. N., Braden, J. P., & White, J. L. (2001). *Assessing one and all: Educational accountability for students with disabilities*. Arlington, VA: Council for Exceptional Children.
- Elliott, J. L., & Thurlow, M. L. (2000). *Improving test performance of students with disabilities...on district and state assessments*. Thousand Oaks, CA: Corwin Press.
- Elliott, J. L., Thurlow, M. L., & Ysseldyke, J. E. (1996). *Assessment guidelines that maximize the participation of students with disabilities in large-scale assessments: Characteristics and considerations* (Synthesis Report No. 25). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved October 23, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis25.html>
- Erickson, R., Ysseldyke, J., Thurlow, M., & Elliott, J. (1998). Inclusive assessment and accountability systems: Tools of the trade in educational reform. *Teaching Exceptional Children, 31*(2), 4-9.
- Ford, A., Davern, L., & Schnorr, R. (2001). Learners with significant disabilities: Curricular relevance in an era of standards-based reform. *Remedial and Special Education, 22*(4), 214-222.
- Fowler, F. C. (2000). *Policy studies for educational leaders: An introduction*. Upper Saddle River, NJ: Prentice-Hall.
- Furney, K. S., Hasazi, S. B., Clark/Keefe, K., & Hartnett, J. (2003). A longitudinal analysis of shifting policy landscapes in special and general education reform. *Exceptional Children, 70*(1), 81-94.

- Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). White Plains, NY: Longman.
- Glatthorn, A. A. (1998). *Writing the winning dissertation: A step-by-step guide*. Thousand Oaks, CA: Corwin Press.
- Goals 2000: Educate America Act, 20 U.S.C. (1994).
- Hager, K. D., & Slocum, T. A. (2002, March). Alternate assessment: No child left behind during statewide testing. In *No child left behind: The vital role of rural schools*. Annual National Conference Proceedings of the American Council on Rural Special Education (ACRES). Reno, NV.
- Haigh, J. (1999). *Accommodations, modifications, and alternates for instruction and assessment* (Maryland/Kentucky Report 5). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved September 24, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/MdKy5.html>
- Heumann, J., & Warlick, K. (2000, August 24). *Questions and answers about provisions in the Individuals with Disabilities Education Act Amendments of 1997 reslated to students with disabilities and state and district-wide assessments* [Memorandum]. Washington, DC: U.S. Department of Education.
- Hill, R. (2001, June). *The impact of including special education students in accountability systems*. Paper presented at the 31st Annual Conference on Large-Scale Assessment, Houston, TX.
- Hill, R. K., & DePascale, C. A. (2003, February). *Reliability of No Child Left Behind accountability designs*. The National Center for the Improvement of Educational Assessment, Inc. Retrieved September 29, 2003 from,

http://www.nciea.org/publications/NCME_RHCD03.pdf

Hoff, D. (2004, March 10). Accountability conflicts vex schools. *Education Week*, p. 1, 23.

Howley-Rowe, C., & Leopold, G. D. (2000). *Evaluation report of the AEL CSRD academy for external facilitators: 2000 Summer Institute*. Available from: AEL, Inc. P.O. Box 1348, Charleston, WV 25325-11348.

Human Resource Development Press. (1995). *Change abilitator: A tool for managing change (Leader's guide)*. Amherst, MA: Author.

The impact of alternate assessment and access to the general curriculum grants update.

(2003, October 11). Retrieved November 15, 2003, from

<http://www.uncc.edu/aap/aafir/aafir.asp>

Improving the Academic Achievement of the Disadvantaged; Final Rule, *Fed. Reg.* 68, 236 (Dec. 9, 2003) (to be codified at 34 C.F.R. p. 200).

Improving America's Schools Act of 1994, 20 U.S.C. (1994).

Individuals with Disabilities Education Act, Amendments of 1997, 20 U.S.C (1997).

Kearns, J. F., Kleinert, H. L., & Kennedy, S. (1999). We need not exclude anyone.

Educational Leadership, 56(6), 33-38.

Kentucky Alternate Portfolio Project. Retrieved November 21, 2003, from

<http://www.ihdi.uky.edu/kap/index.asp>

Kentucky Department of Education. Retrieved November 28, 2003, from

<http://www.education.ky.gov>

Kentucky Department of Education. (2002). *2002-2003 DAC implementation guide*.

Frankfort: Author.

- Kentucky Department of Education. (2003). *2003 CATS interpretive guide*. Frankfort: Author.
- Kleinert, H., Kennedy, S., & Kearns, J. (1999). The impact of alternate assessments: A statewide teacher survey. *Journal of Special Education, 33*(2), 93-102.
- Kleinert, H., & Thurlow, M. (2001). An introduction to alternate assessment. In H.L. Kleinert & J. F. Kearns (Eds.), *Measuring outcomes and supports for students with disabilities* (pp. 1-15). Baltimore: Paul H. Brookes.
- Kleinert, H. L. (2001). An introduction to alternate assessment. *Assessment for Effective Intervention, 26*(2), 57-60.
- Kleinert, H. L., Haigh, J., Kearns, J. F., & Kennedy, S. (2000). Alternate assessments: Lessons learned and roads to be taken. *Exceptional Children, 67*(1), 51-66.
- Kleinert, H. L., & Kearns, J. F. (2001). *Alternate assessment: Measuring outcomes and supports for students with disabilities*. Baltimore: Brookes.
- Krentz, J., Thurlow, M., & Callender, S. (2000). *Accountability systems and counting students with disabilities* (Technical Report 29). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved October 23, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/Technical29.htm>
- Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA : Sage Publications.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Beverly Hills, CA : Sage Publications.
- La Marca, P. M. (2001). Alignment of standards and assessments as an accountability criterion. *Practical Assessment, Research & Evaluation, 7*(21). Retrieved

February 12, 2003, from <http://pareonline.net/getvn.asp?v=7&n=21>

- Linn, R. L. (2001). Assessments and accountability (condensed version). *Practical Assessment, Research & Evaluation*, 7 (11). Retrieved January 30, 2003 from: <http://pareonline.net/getvn.asp?v=7&n=11>
- Linn, R. L., Baker, E. L., & Herman, J. L. (2002, Fall). Minimum group size for measuring adequate yearly progress. *The CRESST Line*, 1, 4-5.
- Maryland State Department of Education. (n.d.). Retrieved November 27, 2003, from www.marylandpublicschools.org/
- Maryland State Department of Education. (2003). *Alternate Maryland school assessment (Alt-MSA) test administration and coordination manual*. Baltimore: Author.
- McDonnell, L. M., McLaughlin, M. J., & Morison, P. (Eds.). (1997). *Educating one and all: Students with disabilities and standards-based reform* [Electronic version]. Washington, DC: National Academy Press.
- McLaughlin, M. J., Henderson, K., & Rhim, L. M. (1998). *Snapshots of reform: How five local districts are interpreting standards-based reform for students with disabilities*. Alexandria, VA: Center for Policy Research on the Impact of General and Special Education Reform.
- Meister, D. G. (2000, April). *Teachers and change: Examining the literature*. Paper presented at the annual meeting of the American Education Research Association. New Orleans, LA.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks: CA, Sage.
- A nation at risk: The imperative for educational reform*. Retrieved January 10, 2003,

from U.S. Department of Education Publications Online at

<http://www.ed.gov/pubs/NatAtRisk>

National Center on Educational Outcomes. (NCEO). (n.d.). Retrieved February 19, 2003,

from <http://education.umn.edu/nceo/>

National Center on Education Outcomes. (2003). *Accountability for assessment results in the No Child Left Behind act: What it means for children with disabilities.*

Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Retrieved October 23, 2003, from

<http://education.umn.edu/nceo/OnlinePubs/NCLBdisabilities.html>

National Research Council. (1999). *Testing, teaching, and learning: A guide for states*

and school districts. Committee on Title I Testing and Assessment, R. F. Elmore

and R. Rothman (Eds.). Board on Testing and Assessment, Commission on

Behavioral and Social Sciences and Education. Washington, DC: National

Academy Press.

Neuendorf, K. A. (2002). *The content analysis guidebook.* Thousand Oaks, CA: Sage.

No Child Left Behind Act of 2001, Pub. L. No 107-110, 115 Stat. 1425 (2002).

North Carolina Department of Public Instruction. (2000, February). *1999-2000 North*

Carolina alternate assessment portfolio pilot. Raleigh: Author.

North Carolina Department of Public Instruction. (2001). *Portfolio development designee*

manual. Raleigh: Author.

North Carolina Department of Public Instruction. (2002, April). *North Carolina alternate*

assessment portfolio scoring model: 2001-2002. Raleigh: Author

North Carolina Department of Public Instruction. (2003). *Portfolio development*

- designee/teacher handbook*. Raleigh: Author.
- North Carolina State Board of Education. (2003). *Testing students with disabilities: North Carolina testing program*. Raleigh: Author.
- Olsen, K. (1999a, May). *Alternate assessment issues and practices*. Lexington: University of Kentucky, Mid-South Regional Resource Center. Retrieved June 19, 2002, from <http://www.ihdi.uky.edu/msrrc/Publications/aaissues.htm>
- Olsen, K. (1999b, May). *What principles are driving development of state alternate assessments?* Lexington: University of Kentucky, Mid-South Regional Resource Center. Retrieved June 19, 2002, from <http://www.ihdi.uky.edu/msrrc/Publications/whatprincip.htm>
- Olson, B., Mead, R., & Payne, D. (2002). *A report of a standard setting method for alternate assessments for students with significant disabilities* (Synthesis Report 47). Minneapolis: University of Minnesota, National Center on Educational Outcomes, Retrieved November 12, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis47.html>
- Paige, R. (2002, July 24). *Dear colleague letter to education officials regarding implementation of the No Child Left Behind Act*. Retrieved January 16, 2003, from <http://www.ed.gov/policy/elsec/guid/secletter/020724.html?exp=0>
- Quenemoen, R. F., Lehr, C. A., Thurlow, M. L., & Massanari, C. B. (2001). *Students with disabilities in standards-based assessment and accountability systems: Emerging issues, strategies, and recommendations* (Synthesis Report 37). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved October 23, 2002, from

<http://education.umn.edu/NCEO/OnlinePubs/Synthesis37.html>

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 September 24, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis43.html>

Quenemoen, R., Thompson, S., Thurlow, M., & Lehr, C. (2001). *A self-study guide to implementation of inclusive assessment and accountability systems: A best practice approach*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved September 24, 2004, from <http://education.umn.edu/nceo/OnlinePubs/workbook.pdf>

Quenemoen, R., Thompson, S., Thurlow, M., & Olsen, K. (1999). *Forum on alternate assessment and "gray area" assessment*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved September 23, 2002, from http://education.umn.edu/NCEO/GrayAreaForum/Conference_Report.htm

Quenemoen, R., & Thurlow, M. (2002). *Including alternate assessment results in accountability decisions* (Policy Directions No. 13). Minneapolis: University of Minnesota, National Center on Education Outcomes. Retrieved October 23, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/Policy13.htm>

Rist, R. C. (2000). Influencing the policy process with qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1001-1017). Thousand Oaks, CA: Sage.

Roeber, E. (2002). *Setting standards on alternate assessments* (Synthesis Report 42).

Minneapolis: University of Minnesota, National Center on Education Outcomes.

Retrieved November 12, 2002, from

<http://education.umn.edu/NCEO/OnlinePubs/Synthesis42.html>

Rudestam, K. E., & Newton, R. R. (2001). *Surviving your dissertation: A comprehensive guide to content and process* (2nd ed.). Thousand Oaks, CA: Sage.

Smith, M. S., & O'Day, J. (1991, April). *Putting the pieces together: Systemic school reform* (Consortium for Policy Research in Education Policy Brief No. RB-06).

Philadelphia: University of Pennsylvania, Graduate School of Education.

Retrieved February 6, 2003, from <http://www.cpre.org/Publications/rb06.pdf>

Spring, J. (1998). *Conflict of interests: The politics of American education* (3rd ed.).

Boston: McGraw-Hill.

Strauss, A., & Corbin, J. (1998). *Basis of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

Tennessee Department of Education. (n.d.) Retrieved November 7, 2003, from

<http://www.state.tn.us/education/index.html>

Tennessee Department of Education. (2002). *TCAP-ALT portfolio manual, 2003-2004*.

Nashville: Author.

Tennessee State Board of Education. (n.d.). Retrieved November 7, 2003, from

<http://www.state.tn.us/sbe/>

Thompson, S., Erickson, R., Thurlow, M., Ysseldyke, J., & Callender, S. (1999). *Status of the states in the development of alternate assessment* (Synthesis Report 31).

Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Retrieved November 3, 2001, from

<http://education.umn.edu/NCEO/OnlinePubs/Synthesis31.html>

- Thompson, S., Quenemoen, R., Thurlow, M., & Ysseldyke, J. (2001). *Alternate assessments for students with disabilities*. Thousand Oaks, CA: Corwin Press.
- Thompson, S., & Thurlow, M. (1999). *1999 State special education outcomes: A report on state activities at the end of the century*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved September 23, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/99StateReport.htm>
- Thompson, S. J., & Thurlow, M. L. (2000). *State alternate assessments: Status as IDEA alternate assessment requirements take effect* (Synthesis Report No. 35). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved October 23, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis35.html>
- Thompson, S., & Thurlow, M. L. (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 October 23, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/2001StateReport.html>
- Thompson, S., & Thurlow, M. (2003a, April). *Effective implementation of alternate assessments for students with severe disabilities*. Paper presented at the annual meeting of the Council for Exceptional Children, Seattle, WA.
- Thompson, S., & Thurlow, M. (2003b). *2003 State special education outcomes: Marching on*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved April 11, 2004, from

<http://education.umn.edu/nceo/>

Thurlow, M. L. (1994). *National and state perspectives on performance assessment and students with disabilities*. Reston, VA: The Council for Exceptional Children.

Thurlow, M. L., Elliott, J. L., & Ysseldyke, J. E. (1998). *Testing students with disabilities: Practical strategies for complying with district and state requirements*. Thousand Oaks, CA: Corwin Press.

Thurlow, M. L., & Johnson, D. R. (2000). High-stakes testing of students with disabilities. *Journal of Teacher Education*, 51(4), 305-314.

Thurlow, M., Olsen, K., Elliott, J., Ysseldyke, J., Erickson, R., & Ahearn, E. (1996). *Alternate assessment for students with disabilities for students unable to participate in general large-scale assessments* (Policy Direction No. 5).

Minneapolis: University of Minnesota, National Center on Education Outcomes.

Retrieved November 12, 2002, from

<http://education.umn.edu/nceo/OnlinePubs/Policy5.html>

Thurlow, M., Quenemoen, R., Thompson, S., & Lehr, C. (2001). *Principles and characteristics of inclusive assessment and accountability systems* (Synthesis Report 40). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved September 24, 2002, from

<http://education.umn.edu/NCEO/OnlinePubs/Synthesis40.html>

Thurlow, M. L., Ysseldyke, J. E., Gutman, S., & Geenan, K. (1998). *An analysis of inclusion of students with disabilities in state standards documents* (Technical Report No. 19). Minneapolis: University of Minnesota, National Center on

- Educational Outcomes. Retrieved November 23, 2001, from <http://education.umn.edu/NCEO/OnlinePubs/Technical19.html>
- Turnbull, A., Turnbull, R., Shank, M., & Leal, D. (1999). *Exceptional lives: Special education in today's schools* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Turner, M. D., Baldwin, L., Kleinert, H., & Kearns, J. F. (2000). The relation of a statewide alternate assessment for students with severe disabilities to other measures of instructional effectiveness. *The Journal of Special Education, 34*(2), 69-76.
- U.S. Department of Education. (1994). *Progress of education in the United States of America - 1990 through 1994*. Washington, DC: Author.
- U.S. Department of Education, Office of Special Education Programs. (2002). *Twenty-fourth annual report to Congress on the implementation of the Individuals with Disabilities Education Act*. Washington, DC: Author.
- U.S. Department of Education. (2004, January 14). *Charting the courses: States decide major provisions under No Child Left Behind*. Washington, DC: Author.
- U.S. General Accounting Office. (1996). *Content analysis: A methodology for structuring and analyzing written material* (Report No. USGAO/PEMD-10.3.1). Washington, DC: U.S. G.A.O. Program Evaluation and Methodology Division.
- Vanderwood, M., McGrew, K. S., & Ysseldyke, J. E. (1998). Why we can't say much about students with disabilities during education reform. *Exceptional Children, 64* (3), 359-370.
- Virginia Department of Education. (n.d.). Retrieved February 12, 2003, from <http://www.pen.k12.va.us>

Virginia Department of Education. (2000a). *Field test manual, March 31, 2000.*

Richmond: Author.

Virginia Department of Education. (2000b). *Field test manual, April 7, 2000.*

Richmond: Author.

Virginia Department of Education. (2000c, March 24). *Participation criteria for alternate assessment, March 24, 2000.* [Supts. Memo. No. 61]. Richmond:

Author.

Virginia Department of Education. (2000d). *Virginia alternate assessment program: Administrators' manual.* Richmond: Author.

Virginia Department of Education. (2001). *Virginia alternate assessment program: Implementation manual.* Richmond: Author.

Virginia Department of Education. (2002). *Virginia alternate assessment program: Implementation manual - Revised 2002.* Richmond: Author.

Warlick, K. R. (2001). Foreword. In H. L. Kleinert & J. F. Kearns. *Alternate assessment: Measuring outcomes and supports for students with disabilities* (pp. xi-xii).

Baltimore: Paul H. Brooks.

Warlick, K., & Olsen, K. (1999, April). *How to conduct alternate assessments: Practices in nine states.* Lexington: University of Kentucky, Mid-South Regional Resource Center. Retrieved October 11, 2002, from <http://www.ihdi.uky.edu/msrrc/Publications/howaa.htm>.

Weber, R. P. (1990). *Basic content analysis.* Newbury Park, CA: Sage Publications.

West Virginia Department of Education. (n.d.). Retrieved November 11, 2003, from <http://wvde.state.wv.us/>

- West Virginia Department of Education. (2003). *West Virginia measures of academic progress: Alternate assessment administration manual*. Charleston: Author.
- Wiener, D. (2002). *Massachusetts: One state's approach to setting performance levels on the alternate assessment* (Synthesis Report 48). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved November 25, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis48.html>
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco: Jossey-Bass Publishers.
- Yanow, D. (2000). *How does a policy mean? Interpreting policy and organizational actions* [Electronic version]. Washington, DC: Georgetown University Press.
- Ysseldyke, J., Krentz, J., Elliott, J., Thurlow, M. L., Erickson, R., & Moore, M. L. (1998). *NCEO framework for educational accountability*. Minneapolis: University of Minnesota, National Center on Education Outcomes. Retrieved February 3, 2003, from <http://education.umn.edu/NCEO/OnlinePubs/Framework/FrameworkText.html>
- Ysseldyke, J. E., & Olsen, K. (1999). Putting alternate assessments in to practice: What to measure and possible sources of data. *Exceptional Children*, 65(2), 175-185.
- Ysseldyke, J. E., Olsen, K. R., & Thurlow, M. L. (1997). Issues and considerations in alternate assessments (Synthesis Report No. 27). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved September 24, 2002, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis27.htm>
- Ysseldyke, J., Thurlow, M., Erickson, R., Gabrys, R., Haigh, J., Trimble, S., & Gong, B.

(1996). *A comparison of state assessment systems in Maryland and Kentucky with a focus on the participation of students with disabilities* (Maryland-Kentucky Report 1). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved December 14, 2003, from http://education.umn.edu/NCEO/OnlinePubs/MDKY_1.html

Ysseldyke, J. E., Thurlow, M. L., Kozleski, E., & Reschly, D. (1998). *Accountability for the results of educating students with disabilities: Assessment conference report on the new assessment provisions of the 1997 amendments to the Individuals with Disabilities Act*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved February 2, 2003, from <http://education.umn.edu/NCEO/OnlinePubs/awgfinal.html>

VITA

Tamra Roberts Cobb

Birthdate: February 27, 1959

Birthplace: Hampton, VA

Education:

1999-2004	College of William & Mary Williamsburg, Virginia Doctor of Philosophy
1981-1983	University of Illinois at Urbana-Champaign Urbana, IL Master of Arts
1977-1981	Hampton University Hampton, Virginia Bachelor of Arts

Experience:

1999-present	Assistant Professor School of Science Department of Communicative Sciences & Disorders Hampton University Hampton, Virginia
1997-1999	Director of Business Development; Eastern Region & Healthcare Consultant The Polaris Group Hingham, Massachusetts
1995-1997	Rehabilitation Program Manager Springdale HealthCare Center Camden, South Carolina
1994-1995	Department Head; Speech Therapy/Audiology The Queen's Medical Center Honolulu, Hawaii
1991 - 1995	Audiologist The Queen's Medical Center Honolulu, Hawaii