


2006

A Study of New Jersey's Assessment and Accountability System for Students with Disabilities and its Impact on Special Education Services

Julie Moul Gardner
Seton Hall University

Follow this and additional works at: <https://scholarship.shu.edu/dissertations>

 Part of the [Accessibility Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), and the [Special Education and Teaching Commons](#)

Recommended Citation

Gardner, Julie Moul, "A Study of New Jersey's Assessment and Accountability System for Students with Disabilities and its Impact on Special Education Services" (2006). *Seton Hall University Dissertations and Theses (ETDs)*. 2348.
<https://scholarship.shu.edu/dissertations/2348>

A STUDY OF NEW JERSEY'S ASSESSMENT AND ACCOUNTABILITY SYSTEM
FOR STUDENTS WITH DISABILITIES AND ITS IMPACT ON SPECIAL
EDUCATION SERVICES

BY

JULIE MOUL GARDNER

Submitted in partial fulfillment of the
requirements for the Degree of Doctor of Education
Seton Hall University
2006

Dissertation Committee

Barbara Strobert, Ed.D., Mentor
Michael Osnato, Ed.D.
Janet L. Parmelee, Ed.D, J.D.
Lloyd M. Leschuk, Ed.D.

Submitted in partial fulfillment of the
requirements for the Degree of Doctor of Education
Seton Hall University
2006

ABSTRACT

A Study of New Jersey's Assessment and Accountability System for Students with Disabilities and its Impact on Special Education Services

Use of high stakes assessment for special needs students is a key element in the current educational reform movement. State assessment systems use these tests to determine and report school accountability scores, distribute school funding, and assess school rewards and sanctions. The impact of including these students' test score on the district accountability reports has not yet been determined. Directors of special services, the primary individuals for supervising or coordinating district special education programs, have an extensive understanding of administrative and functional knowledge of special education services and are therefore able to provide insight concerning the impact of accountability on special education services.

This study collects and describes directors' of special services perceptions on the impact of high-stakes assessment and special education services. Data were gathered in this study through both survey and focus group format.

The findings of this study indicate accountability in special education has improved the instructional process for students with disabilities. Professional development opportunities have improved for general education teachers. The inclusion of students with disabilities is viewed with mixed results. We still do not know the full impact of including students in high-stakes large-scale testing. Systematic concerns related to the reliability and validity of the alternate assessment process exist. Questions remain on whether the assessment meets the desired outcome of measuring student performance.

| | |
|---|-----------|
| TABLE OF CONTENTS | |
| DEDICATION | vii |
| ACKNOWLEDGEMENTS | viii |
| LIST OF TABLES | iii |
| CHAPTER 1: INTRODUCTION | 1 |
| Statement of the Problem | 3 |
| Purpose of the Study | 4 |
| Significance of the Study | 5 |
| Research Questions | 7 |
| Procedures | 9 |
| Limitations of the Study | 10 |
| Definition of Terms | 11 |
| Summary | 14 |
| CHAPTER II: LITERATURE REVIEW | 16 |
| Overview | 16 |
| Legislation | 16 |
| Standards Based Reform | 19 |
| Adequate Yearly Progress | 21 |
| Accountability Through Assessment | 24 |
| Assessment Accommodations | 27 |
| Alternate Assessment | 31 |
| Assessment Validity and Reliability | 36 |
| New Jersey Assessment and Accountability System | 46 |
| Effects of Inclusion in Testing | 49 |
| Summary | 57 |
| CHAPTER III: METHODOLOGY | 58 |
| Overview | 58 |
| Methodology | 59 |
| Research Design | 60 |
| Subjects | 65 |
| PASE Survey Participants | 65 |
| Focus Group Participants | 67 |
| Data Collection | 68 |
| PASE Survey | 68 |
| Focus Group | 69 |
| Instrument Design | 69 |
| PASE Survey design | 69 |

| | |
|--|-----|
| Summary | 95 |
| CHAPTER IV: RESEARCH FINDINGS | 95 |
| Summary | 95 |
| Findings | 95 |
| Implications | 102 |
| Focus Group Design | 71 |
| Data Analysis | 71 |
| Analysis of PASE Survey | 71 |
| Analysis of the Focus Group | 72 |
| Summary | 72 |
| CHAPTER IV: RESEARCH FINDINGS | 73 |
| Overview | 73 |
| Quantitative Research Results | 75 |
| Research Question 1 | 75 |
| Research Question 2 | 76 |
| Hypothesis for Question 2 | 77 |
| Research Question 3 | 79 |
| Hypothesis for Question 3 | 80 |
| Research Question 4 | 82 |
| Hypothesis for Question 4 | 83 |
| Research Question 5 | 85 |
| Research Question 6 | 85 |
| Research Question 7 | 86 |
| Research Question 8 | 87 |
| Summary of Quantitative Research Results | 88 |
| Qualitative Research Results | 88 |
| Summary | 94 |
| CHAPTER V: SUMMARY | 96 |
| Summary | 96 |
| Findings | 99 |
| Implications | 102 |
| Recommendations | 103 |
| Future Research | 103 |
| State and National Policy Implications | 105 |
| Summary | 106 |
| References | 108 |
| Appendix A: Survey Instrument | 119 |
| Appendix B: Focus Group Instrument | 123 |

DEDICATION

This dissertation is dedicated to my husband Tom and my children Ian and Caitlyn. I cannot express enough gratitude to each of them for their personal support and sacrifices during this study.

It is also dedicated to my parents, James and JoAnn Moul for instilling in me the desire to learn and achieve. They have been responsible for giving me the necessary skills to be successful in life.

Lastly, I dedicate this dissertation to the Morris-Union Jointure Commission's superintendent, assistant superintendent and my fellow principals for providing the incentive to pursue my doctoral degree in education administration. It is their commitment to the field of education and special education that inspired me to continue my professional growth.

ACKNOWLEDGEMENTS

Grateful acknowledgement is expressed to the members of my dissertation committee at Seton Hall University. Dr. Barbara Strobert, my mentor professor, provided professional guidance throughout the project. To Dr. Michael Osnato, much appreciation is expressed for his review of the details of the drafts of this research project. To Dr. Lloyd Leschuk for providing me with valuable feedback and to Dr. Janet Parmelee, my sincerest appreciation who brought important educational perspectives as well as the grammatical corrections throughout this process. My thanks to each of my committee members for giving freely of their time and intellect to help me complete this project.

I truly appreciate all of the directors of special services support in this process and the time they took away from their busy jobs to complete the survey and participate in the focus group.

| LIST OF TABLES | | |
|----------------|--|----|
| 1 | Summary of Frequencies for Changes in Student Achievement and Instruction | 75 |
| 2 | Summary of Received Surveys with Respect to DFG | 77 |
| 3 | Descriptive Statistics for Accountability by DFG Group | 77 |
| 4 | Test of Homogeneity of Variances for Accountability by DFG Group | 78 |
| 5 | ANOVA for Accountability by DFG Group | 78 |
| 6 | Summary of Received Surveys with Respect to District Enrollment Groups | 79 |
| 7 | Descriptive Statistics for Accountability by District Enrollment Group | 80 |
| 8 | Test of Homogeneity of Variances for District Enrollment Group | 81 |
| 9 | ANOVA for District Enrollment Group | 81 |
| 10 | Summary of Received Surveys with Respect to Years of Experience | 82 |
| 11 | Descriptive Statistics for Accountability by Years of Experience | 83 |
| 12 | Test of Homogeneity of Variances for Years of Experience | 84 |
| 13 | ANOVA of Years of Experience | 84 |
| 14 | Summary of Frequencies with Respect to Changes in Use of Testing Accommodations | 85 |
| 15 | Summary of Frequencies with Respect to Professional Development | 86 |
| 16 | Summary of Frequencies with Respect to the Effectiveness of the Alternate Proficiency Assessment | 86 |
| 17 | Summary of Frequencies with Respect to the Effectiveness of Inclusion in Testing | 87 |
| 18 | Summary of Focus Group Responses | 92 |

Chapter I

Introduction

The purpose of this study was to explore the consequences of large-scale high-stakes assessment, with specific reference to the New Jersey Alternate Proficiency Assessment results for students with disabilities through the point of view, experiences, and perceptions of the directors of special services in 100 public schools in central and northern New Jersey. The information gathered through the use of a survey provided insights into consequential aspects of large-scale assessment as it is practiced in New Jersey.

Educators, administrators, parents, citizens, and policymakers need to know the extent to which all students are learning and benefiting from the educational programming in local districts. As a means to assess this progress, states have implemented large-scale, high-stakes testing. Improvement in test scores is the primary measure that states use to report to the public whether schools are improving and students are achieving. In the political arena, accountability in education equates to high-stakes testing and disclosure of the test results to the public.

Accountability has become linked to financial assistance, school quality, teacher competency and student progress. Numerous researchers view assessment as a process to motivate teachers to provide quality instruction, motivate students to achieve, measure a student's performance, and ensure quality educational services (Browder, Fallin, Davis, & Karvonen (2003), Kampfer, Horvath, Kleinert, & Kearns (2001), Thompson,

Quenemoen, Thurlow, Ysseldyke (2003), Thompson, & Thurlow, (2001), Turner, Baldwin, Kleinert, & Kearns (2001). However, conflicting evidence is emerging that questions the impact of high-stakes testing, especially with regard to improved student learning (Amerein & Berliner, 2002). Additional questions have focused on the effects on the curriculum, student and teacher attitudes, school climate, and the financial impact of testing. (Elliott, Erickson, Thurlow, & Shriner, 2000).

No Child Left Behind Act of 2001 (NCLB 2001) and the quest for improved schools have elevated the importance of assessment results for the purpose of school improvement and student achievement for all students, including students with disabilities. Schools can no longer ignore the challenges faced by students with disabilities and exclude them from standardized statewide and district-wide testing in reading, math and science.

State and district reports must aggregate the data as well as disaggregate the data by economic, ethnic, disability, gender and English-proficiency status (NCLB 2001, Title I, Part A, Section 1111a2C, 2001). According to the U.S. Department of Education (Facts Page on Adequate Yearly Progress, 2003), the purpose of disaggregating the data is to assure that "failure cannot hide." The goal of NCLB 2001 is to close the achievement gaps between students of different groups by, "holding schools accountable for the achievement of all subgroups," thereby assuring that "no child is left behind" (U.S. Department of Education, Facts Page on Adequate Yearly Progress, 2003). Because these results are readily available in state reports, the public, media, policymakers, and educational researchers are able to use the results to interpret performance levels and performance trends for students with disabilities.

While the idea of students with disabilities being included in a standards-based reform movement is welcomed by educators and parents, it remains unclear how to accurately and appropriately assess special education students, how to assure the students can achieve the same educational standards as other students and the ultimate impact of reporting test results of special education students on both the schools' and districts' adequate yearly progress indices (Albrecht & Joles (2003); Cohen & Heumann, (2001).

Statement of the Problem

This study will employ quantitative and qualitative research to explore the consequences of large-scale high-stakes assessment and its impact on the increased demand for accountability on special education services in 100 public school districts in central and northern New Jersey.

With the passage of NCLB 2001, educators, for the first time, must confront what it really means for all students to achieve high standards, including students with the most severe disabilities. Students with disabilities have historically been exempt from high-stakes testing with limited expectations of accountability for performance. This legislation indicates a policy shift away from a special education system grounded in meeting individualized goals to inclusion in a standardized system of accountability. The focus in special education has changed from concerns regarding inclusion of students into regular education classrooms to inclusion of these students in the standardized assessment process.

With this tremendous task facing educators, a central question remains how to accurately and appropriately assess this diverse group and the impact of reporting the

results on a school district's adequate yearly progress. Coupled with this diversity is the increasing number and percentage of students receiving special education services both in New Jersey and nationally. According to the *Quality Counts 2004 Report* (Education Week, 2004), special education services are increasing at a rate six times faster than the regular education population. The combination of an increasing number of special education students and the stringent accountability requirements dictates a need to explore the consequences of large-scale high-stakes assessment for students with disabilities.

Purpose of the Study

The purpose of this study was to explore the consequences of large-scale high-stakes assessment, with specific reference to the New Jersey Alternate Proficiency Assessment results for students with disabilities through the point of view, experiences, and perceptions of the directors of special services in 100 public schools in central and northern New Jersey. The information gathered through use of a survey will provide insights into consequential aspects of large-scale assessment as it is practiced in New Jersey. Questions will focus on student performance, accommodations, alternate assessments, accountability, and current and emerging practices and issues. The focus group venue is designed to provide insight and experiences from the participants into the consequential aspects of large-scale assessment and the impact on accountability of special education services as it is practiced in New Jersey.

Significance of the Study

The inclusion of students with disabilities in large-scale high-stakes testing is mandated by federal and state legislation. As reported by the National Center for Educational Outcomes (Thurlow, Wiley & Bielinski, 2003), special education advocates support the inclusion in the assessment process based on the ideal that the information provided by testing will result in improved educational opportunities for students with disabilities. This trend holds both promise and peril for these students. States have been making progress to include performance scores of special education students in the district adequate yearly progress as required by NCLB 2001. According to Education Week *Quality Counts 2004* report (2004, p.7), 43 states and the District of Columbia include information on the test participation rates and performance of students with disabilities in the reporting system.

The Council for Exceptional Children (Allbritten, Mainzer, & Zeigler 2004) reports that despite agreement that students with disabilities should be included in testing and receive the benefits of the test outcomes, questions concerning the reliability and validity of test results when accommodations and alternate assessments are used, the accuracy of reporting those results, and ultimately the impact on special education services of including the results in the state's report card remain.

Individuals with Disabilities Act, 2004 and NCLB 2001, require students with disabilities to receive appropriate accommodations and modifications for standardized assessments as determined by the individualized education program (IEP) team. However, the use and types of accommodations are controversial and raise questions relating to the validity of the standardized assessment when accommodations are allowed.

The IEP team is also given the responsibility to determine if a student should be assessed through the use of an alternative assessment, when even with accommodations, the student is unable to take the standardized test. Prior to the regulation change on December 9, 2003 (U.S. Department of Education, 2003), districts were required to report all scores, even scores of assessments taken by students with the most significant cognitive disabilities, based on the same grade-level standards. These scores were frequently scored as non-proficient, thus hindering state and district efforts to reach adequate yearly progress as defined in NCLB 2001.

Under the final rules published in the December 9, 2003 Federal Register, states are permitted to establish alternative assessment standards for students with the most significant cognitive disabilities for the purpose of satisfying the adequate yearly progress requirement of NCLB 2001. The alternate academic achievement standards must clearly define the connection between the instructional content appropriate for non-disabled students and the related knowledge and skills that may serve as a basis for definition of proficient achievement for students with the most significant cognitive disabilities (Special Education Report, 2004, p. 6).

States are permitted to use alternate achievement standards to evaluate the performance of students with the most significant cognitive disabilities and to give equal weight to proficient and advanced performance based on the alternate standards in calculating adequate yearly progress, provided the number of proficient and advanced scores based on the alternate achievement standards does not exceed 1 percent of all students in the grades tested at the state or district level. (Special Education Report, 2004, p. 7).

Even with these changes in NCLB 2001, states may not exclude the scores of students who exceed the percentage cap from the adequate yearly progress calculations. States must count the non-proficient scores for these students and for whom no exception is

granted (Special Education Report, 2004, p. 8). There is no limit to the number of students who take an alternate assessment, just the percentage of scores based on alternate assessment standards that can be counted as proficient when calculating adequate yearly progress.

Although states have some flexibility in reporting the scores of students with the most significant cognitive disabilities, districts are still required to include these students' scores as well as all students with disabilities scores on school and district report cards. This change makes an allowance, and thus an explanation for minimal progress, for the reporting of the scores for students with the most significant cognitive disabilities. However, the impact of the scores for students not meeting the 1% allowance on district adequate yearly progress is still unresolved. Even though the federal government has addressed the reporting of scores, questions remain as to the validity of the alternate assessments and the accommodations permitted when testing students that do not fall within the 1% exception. Therefore, it seems vital to explore the consequences of the use of accommodations and alternate assessments on large-scale high-stakes test results for students with disabilities and the impact of reporting these results on special education services.

Research Questions

On the basis of the previous research and related literature, the research questions were designed to provide insights into consequential aspects of large-scale assessment as it is practiced in New Jersey by answering the following questions:

1. What is the perception of the directors of special services toward changes in student achievement and instruction of special education students since the inclusion of students with disabilities in the New Jersey state assessment system?
2. What is the perception of accountability in special education of directors of special services in low socioeconomic districts (DFG Group) as compared to the perception of accountability in special education of directors of special services in high socio-economic districts (DFG Group)?
3. What is the perception of accountability in special education of director's of special services in high-enrollment districts as compared to the perception of accountability in special education of directors of special services in low-enrollment districts?
4. What is the perception of accountability in special education of directors of special services with less experience as compared to the perception of accountability in special education of directors of special services with more experience?
5. What is the of perception of the directors of special services toward changes in use of testing accommodations since the inclusion of students with disabilities in the New Jersey State assessment system?
6. What is the diversity of perceptions of the directors of special services toward changes in professional development since the inclusion of students with disabilities in the New Jersey State assessment system?

7. What is the diversity of perceptions of the directors of special services toward the effectiveness of the Alternate Proficiency Assessment since the implementation of New Jersey's State alternative assessment system for students with severe disabilities?

8. What is the diversity of perceptions of the directors of special services toward the effect of including students with disabilities in the New Jersey statewide assessment program?

Procedures

This descriptive study employed two methodologies: administering a survey Perceptions of Accountability in Special Education (quantitative design) and conducting a focus group session (qualitative design), in an attempt to explore directors of special services' perceptions regarding the consequences of high-stakes assessment for students with disabilities and the impact of accountability on special education services.

Data were assembled from the survey and the focus group session for analysis. Quantitative analysis procedures included basic descriptive statistics and an analysis of variance (ANOVA) to examine differences between the participant groups involved in the study. Qualitative analysis included coding of the director of special services comments according to themes and patterns.

Limitations of the Study

This investigation was limited by the following elements:

1. New Jersey's assessment and accountability is unique compared with other states. Information collected from this study may not be generalized to other large-scale assessments.
2. Participants' responses were representative of their individual experiences and past and current job responsibilities.
3. This study was not designed to show a causal relationship between the consequences of large-scale testing and educational outcomes for students with disabilities.
4. Differences in populations, practices, and policies in the 100 school districts in central and northern New Jersey may lead to different findings with regard to the questions addressed in this study.
5. Abbott districts with an enrollment of greater than 9000 were not included in the study. This may lead to different results with regard to the DFG socioeconomic group findings.

De finition of Terms

The following selected terms are defined as they were used throughout this study:

Accommodations: Changes made to the assessment procedures in order to provide a student access to information and an equal opportunity to demonstrate knowledge and skills. Accommodations should be chosen based on the IEP team's recommendations and reflect accommodations currently used in daily instruction.

Accountability: Includes schools and districts being accountable for student outcomes by reviewing, evaluating, and improvement in student achievement. Due to NCLB 2001, accountability is linked to rewards and sanctions.

Abbott District: Refers to the New Jersey school districts that receive state aid that is calculated to provide them with the same per-pupil operating budget as would be found in New Jersey's wealthiest school districts. "Abbott" is the short-hand description of a series of New Jersey Supreme Court decisions growing out of litigation filed in 1981 on behalf of children residing in New Jersey's most economically disadvantaged municipalities.

Adequate Yearly Progress (A YP): Refers to the growth rate in the percentage of students who achieve the states definition of academic proficiency. Each state will set the AYP gains every school must meet to reach 100% proficiency at the end of 12 years. Schools are held accountable for the achievement of all students, not just average student performance.

Aggregation: The accumulation of a like data into a larger set. For example, the computation of individual school test scores into a district's score.

Alternate assessment: A method of measuring the performance of students who are unable to participate in statewide assessments, even with accommodations.

Core Curriculum Content Standards: New Jersey's descriptions of the knowledge and skills students should acquire in a particular subject area. New Jersey includes three content areas in the state-wide assessment process: language arts, mathematics, and science

Criterion Referenced Tests: A test that allows its users to make score interpretations in relation to a functional performance level.

Disaggregate: Test results are analyzed by economic, ethnic, disability, gender, and English-proficiency status.

Grade Eight Proficiency Assessment (GEPA): New Jersey eighth grade assessment to determine whether students are making sufficient progress in mastering the knowledge and skills they will need to pass the 11th grade assessment.

High School Proficiency Assessment (HSPA): New Jersey 11th grade assessment used to determine student achievement in reading, writing, and mathematics as specified by the New Jersey Core Curriculum Content Standards.

High-stakes Assessment: An assessment system for which the results have important consequences for students, teachers, and school districts. Examples of high-stakes include promotion, certification, graduation, or school and district rewards or sanctions.

IDEA 2004: An amendment to the Individuals with Disabilities Act which mandates that all students, including students with disabilities, be included in state and district-wide assessments or receive alternate assessments.

Individualized Education Plan (IEP): The individual program plan for the delivery of services to each student with a disability. The intent of the IEP is to assure the student receives an appropriate education in a least restrictive environment.

Large-scale Assessment: Assessments for purposes outside the classroom. Large-scale assessments are usually standardized and give to large groups of students at the same time. Purposes for large-scale assessments include accountability, program evaluation, selection of students for special programs.

Modifications: Changes made to the assessment procedures in order to allow a student to participate. Modifications used during statewide testing should be consistent with instructional assessment procedures used in the classroom.

New Jersey Alternative Proficiency Assessment (NJAPA): Alternate assessment is a portfolio assessment designed to measure progress toward achieving New Jersey's state educational standards for those students with severe disabilities who are unable to participate in the NJASK, GEPA, HSPA

New Jersey Assessment of Skills and Knowledge (NJASK): New Jersey comprehensive, multi-grade assessment program for elementary school age students. The results of the assessment are intended to identify students who need additional instructional support in order to reach the state core curriculum content standards. Replaced the ESPA assessment.

Out-of-level Testing: Administering a test that is designed primarily for people of an age or grade level above or below that of the test taker.

Standards-based Reform: Educational reform in which curriculum is based on content that is clearly defined and meets explicit levels of performance.

Special Review Assessment (SRA): New Jersey alternate assessment for high school students that provides the student with the opportunity to exhibit their understanding and mastery of the HSPA skills in contexts that are familiar and related to their experiences.

Summary

Use of high-stakes assessment for special needs students is a key element in the current educational reform movement. State assessment systems use these tests to determine and report school accountability scores, distribute school funding, and assess school rewards and sanctions. In response to IDEA 2004 and NCLB 2001 student with disabilities are no longer exempt from participation in the high-stakes assessments. However, the impact of including these students' test score on the district accountability reports has not yet been determined. Directors of special services, the primary individuals responsible for supervising or coordinating district special education programs, have an extensive understanding of administrative and functional knowledge of special education services and are therefore able to provide insight into the consequences of high-stakes assessment for students with disabilities, and the impact of accountability on special education services. This study collects and describes directors' of special services perceptions of the impact of high-stakes assessment and special education services. Chapter 2 presents a review of the current review of the literature of high-stakes assessment for students with disabilities. The chapter discusses governmental and legislative impacts on accountability and assessments in special education, standards-based reform, accountability and adequate yearly progress, and alternate assessment and testing accommodations. Chapter 3 describes the setting in

which the research takes place, the data collection process through use of the Perceptions of Accountability in Special Education (PASE) survey (quantitative design) and the focus group (qualitative) and the methodologies that were employed. Chapter 4 summarizes the results of the data analysis for the survey questions and the focus group responses to find out about the directors' of special services perceptions on the impact of high-stakes assessment and special education services. Chapter 5 discusses the research findings and the implications for future research and policy initiatives in the realm of special education services.

Chapter II

Literature Review

Overview

This chapter reviews the literature relating to the topic of assessment and its impact on the increased demand for accountability on special education services. Among the topics discussed in this chapter are: governmental and legislative impacts on accountability and assessments in special education, standards-based reform, accountability and adequate yearly progress, and alternate assessment and testing accommodations.

Legislation

Elementary and Secondary Education Act (ESEA) was reauthorized in 1994 as the Improving America's Schools Act (IASA). IASA required that in Title I funded schools, disadvantaged children served by Title I would meet the same challenging standards for student achievement as expected of other children. All students in the school are held to the same standards, and the progress of all students is measured by these assessments and reported to the public. Based on the assessment reports, schools would make instructional and structural changes needed so the expectations of all students are raised providing opportunities to work toward high standards.

The IASA, the Individuals with Disabilities Education Act (IDEA '97), IDEA 2004 and the No Child Left Behind Act (NCLB 2001 2001) are the most recent statutes to impact the educational reform movement that considers the learning needs of all students,

including those with disabilities. Each of these laws has the common requirement that students with disabilities be included in the assessment system.

IDEA '97 first identified an alternate assessment option for students who cannot participate in general assessments with accommodations. The law implicitly defines state and local assessments as contributing to a student's educational opportunities for which access must be granted. The 1997 IDEA Amendments require that states establish performance goals and indicators for students with disabilities and report student progress to the public (34 CFR 300.138).

On December 3, 2004, IDEA 2004 was signed into law. Specific to this legislation is the expansion of the requirements and definitions of assessment and alternate assessments. The new requirements in IDEA 2004 for explicit participation in the general curriculum and inclusive educational assessments are closely tied to the model for inclusion. IDEA 2004 states:

All children with disabilities be included in all general state and districtwide assessment programs, including assessments described under section 1111 of the ESEA of 1965, with appropriate accommodations and alternate assessments where necessary and as indicated in their respective individualized education program (612 [16] [A]).

IDEA 2004 requires states or local education agencies to develop guidelines for the provision of appropriate accommodations (612 [16] [B]) and further clarifies alternate assessments. According to IDEA 612 [16] [C] [i], the state or local education agency has developed or implemented guidelines for the participation of children with disabilities in alternate assessments. IDEA 2004 includes specific requirements for alternate assessments in that the assessments are aligned with the state's challenging academic

content standards and challenging student academic achievement standards (612 [16] [C] [ii] [I]).

NCLB 2001 requires that no less than 95% of all students in a school participate in state assessments (1111[2]iii). The requirement is broken down into subgroups, with the requirement that 95% of each subgroup must be tested. Under the Act's accountability provisions, states must describe how they will close the achievement gap and make sure all students, including those who are disadvantaged, achieve academic proficiency. They must produce annual state and school district report cards that inform parents and communities about state and school progress. Reporting annual achievement of all students with regard to academic standards enables teachers and parents to identify areas of need and refine instructional programs. Therefore, assuring all students are given the opportunity to receive instruction related to the state's standard performance standards. Schools that do not make progress must provide supplemental services, such as free tutoring or after-school assistance, take corrective actions and, if still not making adequate yearly progress after 5 years, make dramatic changes to the way the school is run (U.S.D.O.E.).

There is no mandate specifying the number or percentage of students with disabilities in states who must participate in standard state assessments. The United States Department of Education (USDOE) (Heumann & Warlick, 2000, p. 8) states, "It is expected that the number of students participating in alternate assessments will be relatively small." NCLB 2001 allows the use of alternate assessments for students with the most severe cognitive disabilities, provided that does not exceed 1% of all students assessed and that the alternate assessment process is aligned with the state's academic

Standards-based reform promotes the setting of standards, identifying indicators of content standards. The assessment must reflect professional judgment of the highest learning possible for those students. The majority of students with disabilities should be participating in standard statewide assessments with appropriate accommodations as needed.

Standards-based Reform

Accountability is a critical aspect of standards-based reform. The rationale for accountability systems is the belief that education can be improved when clear standards for student achievement are communicated to students and educators, achievement toward those standards is measured, and appropriate consequences are linked to levels of student achievement (Lehr & Thurlow, 2003).

Standards-based reform promotes the setting of standards, identifying indicators of how to successfully meet those standards, and ways to measure student progress toward these indicators. The emphasis of standards-based reform is on students achieving high academic standards and that the achievement is documented through testing and assessment (Thurlow, et. al., 2003). The standards-based educational reform movement is intended to produce improvements in instruction and curriculum. Thompson et. al. (2001) defines standards-based reform as including these assumptions:

1. All children are expected to work toward state or district high standards;
2. States and districts measure all children's progress toward the standards.
3. Schools use measurement data to make improvements in curriculum and

insure action to allow all students is to succeed; and

4. States, districts, and schools are held accountable for student success.

Accountability through high-stakes assessment has become a vital component of the educational process with rewards and sanctions for school improvement at the state and local level.

As part of the standards-based reform effort inherent in NCLB 2001, states have developed educational standards along with statewide assessment systems that measure student progress in the core content areas and serve as a basis for district accountability. Content standards are educational standards that define what students are expected to know and to be able to do. They do not guide day to day instruction and are not considered instructional curricula. The standards are skills and abilities needed by all students including the development of basic content area knowledge and understanding, the ability to transfer knowledge to new and different settings, and the ability to apply the learned skills to career or vocational paths, citizenship, or technology (Thompson, et. al, 2001).

Although all students are expected to know and work toward the same high standards, the methods of assessment and the performance expectations among student groups may differ. To assist educators in assessing students appropriately, states have developed performance standards, which are used to describe the quality of the performance expected for proficiency on the content standard. Benchmarks and performance indicators are used to further define measurable and observable skills expected at each grade level.

Even though students will vary in their performance levels, they are expected to learn the same high quality content using a wide range of instructional strategies and methods. According to guidance in the IASA on standards, assessments, and accountability, students with disabilities may need modification to ensure appropriate instructional approaches, but generally all students need to be working toward the same challenging standards (IASA, as cited in Thompson, et. al, 2001, p. 20).

States differ on the design of the content standards, performance indicators and progress indicators. To ensure the inclusion of all students, some states have chosen to expand their content standards to include fundamental skills as performance indicators, but maintaining the same standards for all students while others have separate standards or no common standards for students with severe disabilities. Other states offer examples of skills and performance indicators and then invite IEP teams to select a student's performance indicators from the examples (Thompson, et. al, 2001). However, these differences between states do not lend themselves to easy comparison of data among states for accountability purposes.

Adequate Yearly Progress

Assessment of student progress is a vital component of educational reform. However, prior to the year 2002 state assessment systems were not tied to sanctions or rewards and were viewed as a process not as a means for improvement.

On January 8, 2002, NCLB 2001 was signed in to law (U.S. Department of Education, 2003, Welcome Letter). With the passage of this legislation, the accountability movement in education reached a new height.

The Education Commission of the States (ECS) defined accountability as, "systematic collection, analysis, and use of information to hold schools, educators, and others responsible for the performance of students and the education system" (ECS, 1999 as cited in Thompson et al. 2003, p. 7). Accountability is simply evidence that schools are doing what they are supposed to be doing.

Federal policy and the quest for improved schools has elevated the importance of assessment results for the purpose of school improvement, student achievement, and accountability. NCLB 2001 legislation contains a comprehensive requirement of accountability for states and local education agencies to report annual achievement of all students with regard to academic standards. Among the accountability requirements of NCLB 2001 is that each state must develop and implement a statewide accountability system that ensures all local school districts, and public elementary and secondary schools, make adequate yearly progress (AYP) on annual state assessments in reading and math with science being added by 2005. Adequate yearly progress applies the same high standards of academic performance to all public school students and measures progress of the public schools and local districts on state academic assessments. Standards must be defined by states to include measurable annual goals and objectives for all students and for subgroups, including students with disabilities. Adequate yearly progress must include a timeline for ensuring that each group of students meets or

exceeds a state determined “proficient” level of performance on the state assessment.

NCLB 2001 requires states to:

provide for (I) the participation in such assessments of all students; and (II) the reasonable adaptations and accommodations for students with disabilities as defined under section 602(3) of IDEA '97 as necessary to measure the academic achievement of such students relative to State academic content and State student academic achievement standards.” (1111 [b]3six)

States must provide assessment reports that teachers and parents can use to identify student needs and achievement in relation to state academic standards (NCLB 2001).

The assessments must also, “enable results to be disaggregated within each state, local education agency and school by students with disabilities as compared to students without disabilities” (1111 [b]3). Assessment is viewed as a process to motivate teachers to provide quality instruction, motivate students to achieve, measure a student’s performance, resulting in quality educational services. However, conflicting evidence is emerging that questions the impact of high-stakes testing, especially with regard to improved student learning (Amerein & Berliner, 2002). Additional questions have focused on the effects on the curriculum, student and teacher attitudes, school climate and the financial impact of testing.

Sections 1111(b)(2)(G) and (I) of NCLB 2001 state that schools must show each demographic group is making progress in meeting state objectives and that at least 95 % of the students in each group took the state assessment. According to Section 1111(b)(2)(c)(v), schools do not have to disaggregate data for a demographic group if the number of students is too small to yield statistically reliable information, or the results would reveal personally identifiable information about an individual student.

Accountability Through Assessment

Prior to the implementation of NCLB 2001, students with disabilities were not required to participate in state assessments if their IEP exempt them from the state assessment process. The IDEA 2004 Amendments required participation by all students with disabilities and for states to report the performance of these students at the same rate as for students without disabilities. As a result of NCLB 2001 and IDEA 2004 legislation, states and local education agencies are now held accountable for assessing these students and reporting the test results.

States and local education agencies developed measurements for systems accountability and student accountability to meet the federal mandates. System accountability is designed to improve educational programs, whereas student accountability is designed to motivate students to do their best. Both types of accountability can have unintended negative consequences as well as the intended positive consequences. Thurlow, Elliot, Ysseldyke (1999) state that "A system is accountable for all students when it makes sure that all students count in the evaluation program of the education systems." NCLB 2001 requires that:

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 (1111[b]1).

As the general education program has moved from examining content and process to measuring educational results, such as scores on state-wide assessments, so must the

evaluation of programs for students with disabilities. Thurlow, Elliott, & Ysseldyke, (1999) list five reasons for including students with disabilities in accountability systems.

1. For an accurate picture of education;
2. For students with disabilities to benefit from reforms;
3. To make accurate comparisons;
4. To avoid unintended consequences of exclusion; and
5. To meet legal requirements

McDonnell et al. (1997) list two goals in the efforts to include students with disabilities in assessment and accountability systems. One is to improve the quality of educational opportunities afforded students with disabilities. The second is that inclusion in the assessment process will provide meaningful and useful information about performance of students with disabilities and about the schools that educate them.

The most basic and visible form of accountability is the public reporting of student performance through assessment (EPRRI's Policy Update, 2002). Assessment as defined by Salvia & Ysseldyke (2001) refers to the collection of data thus providing the evidence for accountability. Therefore, it becomes imperative to examine whether current state assessments measure students with and without disabilities equally. State assessments are designed to determine if students have met identified standards of performance related to state content standards. In order to apply the same standard of performance to students with disabilities, the criterion to measure performance must be equally valid for both groups. Since test items are often linked to specific skills from a state's content

standards, test items must measure those skills in the same way for all students in order to make assumptions about student performance based on broad standards.

One of the greatest challenges for states has been the inclusion of more students with disabilities in state testing and accountability systems. Previous to the NCLB 2001 mandate variation of the rates of exclusion of students with disabilities ranged from 33% to 87% (Thurlow, et al., 2003, p. 5). As a result of NCLB 2001, Education Week *Quality Counts 2004* report shows 13 of 37 states provide participation rates for students with disabilities with 95% or more of the students being tested in reading and math for Grades 4, 8, and 10. Participation rates for students with disabilities ranged from 40 to 100%. Twenty of thirty-one states provided participation rates for general education students with larger proportions of general education students tending to take the tests.

Differences in participation rates may reflect the fact that all states do not calculate students the same way when calculating data. All states and the District of Columbia count special education students who take the tests without accommodations, or with accommodations in their participation rates. However, only 26 states count special education students who take state tests with modifications that may result in discrepancies in what is being measured. In the 2002-2003 reporting period, of the 18 states that allow alternative assessments such as portfolio assessment or out of level tests, only 14 states included these results in the reporting of the overall state results (Thurlow & Minnema, 2001).

While more students with disabilities are participating in statewide assessments with and without accommodations, it should be determined whether assessments are benefiting all students the same way. As of 2003, only five states indicated that they

analyzed the assessment results by item for students with disabilities while five states indicated a proposed plan to analyze test items. Fifty percent of the state department of special education directors responded they did not know whether test results were analyzed (Thompson & Thurlow, 2001). These results indicate further discussion between states' assessment agencies and state special education directors is needed to assure test items are assessing student progress or program quality. Eighty percent of state directors of special education reported being involved in their state's development of the definition for adequate yearly progress. (Thompson & Thurlow, 2001)

Assessment Accommodations

To ensure testing of all students, states have defined three options for students to participate in the assessment system: participation in the general assessment, typically a large-scale assessment of some type; participation in the general assessment with accommodations to allow students to show what they know; or participation in alternate ways of assessing what students know and are able to do, commonly called alternate assessment.

Students with disabilities may require accommodations and/or modifications in order to participate in the testing process. Section 504 of the Rehabilitation Act of 1973, the Americans with Disabilities Act of 1990 (ADA), IDEA '97, IDEA 2004 and the NCLB 2001 call for accommodations to be provided as necessary to allow students with disabilities to participate in assessments.

Accommodations are alterations in the way a test is administered, without changing the actual test content or performance standard. An assessment accommodation is provided because of a student need, not to give students with disabilities advantage. When students with disabilities use assessment accommodations it is to show what they know without being impeded by their disabilities. Accommodations offset the impact of the disability and are legally required under Section 504 of the Rehabilitation Act, IDEA '97 (Thompson et al., 2001, p. 49). IDEA 2004 requires states or local education agencies to develop guidelines for the provision of appropriate accommodations.

IDEA 2004 states that the state (or in the case of districtwide assessment, the local education agency) must develop guidelines for the provision of appropriate accommodations (612 [16] [B]). IDEA 2004 does not distinguish between accommodations and modifications in terms of definition of the appropriateness of their use during the administration of statewide assessments. The use of the terms interchangeably in IDEA has caused confusion in the development of consistent policies for the use of accommodations for standardized assessments. Most states have policies listing acceptable accommodations/modifications allowed during state and district wide testing (Thurlow, Scott & Ysseldyke, 1995); however there is inconsistency among the states' allowances. Thurlow, Seyfarth, Scott & Ysseldyke, (1997) indicate that 39 of the 50 states had policies dictating specific accommodations that would be allowed during testing. Most states allow accommodations but restrict their use on portions of the test.

Modifications are commonly defined as changes in setting, timing, scheduling, presentation or response that are considered to alter what a test is supposed to measure. (Thurlow & Weiner, 2000). Modifications can be in the way assessments are designed or

administered so that students with disabilities and limited English proficient students can be included in the assessment. Assessment accommodations or adaptations might include Braille forms for blind students or tests in native languages for students whose primary language is other than English (CRESST).

The New Jersey Department of Education defines accommodations and modifications as changes in how a test is presented to a student, or how a student responds on tests, needed in order for a student with a disability to participate in the test

(N.J. D.O.E.) Accommodations must be listed in the IEP and Section 504 plans. It is important the accommodations used during statewide testing are consistent with instructional assessment procedures used in the classroom (N.J.D.O.E.). Any accommodations must be recorded in the student's test booklet or the answer sheet.

Approved accommodations in the State of New Jersey include: setting, scheduling, test material modifications, and test procedures accommodation and modifications (N.J.D.O.E).

According to Thurlow and Bolt (2001), 34 states allow reading test items aloud with or without limitations although four of those states limit reading aloud to certain portions of the test. In states where reading aloud is an acceptable accommodation, it is not permitted when the test is intended to measure decoding ability and comprehension but may be permitted when the test is intended to only measure comprehension.

In attempting to meet federal mandates for testing requirements states are caught between conflicting ideologies. The politics of accountability are based on all students achieving at the same high level. Students with disabilities requiring instruction and assessment at levels significantly below that of their grade level peers may limit their

meaningful participation in grade level assessments because very few, if any, items on the test would be at an appropriate level of difficulty for these students (Bielinski, Thurlow, Minnema, & Scott, 2000; Gersten & Baker, 2002).

However, the provision of special education services is based on an understanding of a student's individual difference and needs, thus creating the dilemma for educators. The impact of specific accommodations on a test is a factor in the discrepancies among state policies on acceptable accommodations. Research on accommodations is growing rapidly due to the impact of federal legislation on testing requirements. According to Thurlow & Bolt (2001), accommodations for students with physical or sensory disabilities are routinely approved; however, students with cognitive or behavioral difficulties may not receive the modifications necessary due to a lack of research findings on appropriate testing accommodations. Typical accommodations allowed in the classroom such as using a spellchecker mechanism, expressing the student's knowledge and understanding of a question or concept, or demonstrating listening comprehension are not always permitted as accommodations and modifications in high-stakes testing. Cohen and Huemann (2001) found that providing accommodations will improve the test results of students with disabilities but will not interfere with students without disabilities. Cohen and Huemann clarify the responsibility of the IEP team to exercise authority under IDEA '97 to select individual accommodations and modifications necessary for a student with the disability to participate in state and district assessment. Although state and local education agencies cannot limit the authority of the IEP team to select the accommodations and modifications the results of the test may be deemed invalid under

state or local policies if particular accommodations and modifications are provided but not approved by the state and local agencies as per IDEA 2004.

Alternate Assessments

IDEA '97 required states to have an alternate assessment in place by July 1, 2000 (612 [17] [A] [ii]). IDEA 2004 further clarifies alternate assessments. According to IDEA 2004 (612 [16] [C] [i]), the state or local education agency has developed or implemented guidelines for the participation of children with disabilities in alternate assessments. IDEA 2004 includes specific requirements for alternate assessments in that the assessments are aligned with the state's challenging academic content standards and challenging student academic achievement standards (612 [16] [C] [ii] [I]).

Alternate assessments have evolved over the past several years since they were first required by the 1997 IDEA Amendments. Thompson and Thurlow (2001) report that 75% of the states have at least one alternate assessment option that addresses the needs of students with significant cognitive disabilities. The remaining states have two or more alternate assessment options to address the perceived needs of a range of students in addition to those with significant cognitive disabilities.

Alternate assessments are not a traditional large-scale assessment or an individualized assessment. They are designed to measure performance of students who cannot, even with accommodations, participate in general assessment. Perhaps 1% to 2% of the total population of students typically those with very significant disabilities need alternate assessments (Thompson et.al, 2003, p. 8). Using alternative assessment information,

schools can make broad policy decisions that improve schooling practices assuring all students are successful (Thompson et.al, 2003, p. 9).

As a result of the legislative requirement, states have had to rethink the process used to obtain data on student progress for students with disabilities and how to administer an alternate assessment. Several issues emerged around the “what“ of alternate assessment. These issues relate to the alignment of the assessment to the content and performance standards. Students with severe disabilities are often in a curriculum that differs in emphasis from the one that is the course of study for other students. Therefore, the typical test, designed to measure the progress and performance of students in a standard curriculum, often will not be aligned with the curriculum in which students with disabilities receive instruction. Since performance on alternate assessments must be reported with other scores in the assessment system, they must be linked to the high standards that are expected for all students.

Alternate assessments are often viewed as a compromise solution, allowing students to participate in the assessment process while acknowledging their special academic needs. Research-based guidance on successful alternate assessments is limited due to the recent need for an alternate assessment protocol. As a result, the appropriateness of this type of assessment to determine student achievement continues to be reviewed and researched. Alternative approaches are needed to measure the progress of these students toward important educational outcomes. The challenge is that there is a small group of students with severe cognitive deficits or multiple disabilities for whom standard large-scale testing practices and accommodations just do not work.

Arguments supporting alternate assessments include the contention that alternate assessments reduce the frustration and random guessing by student who would otherwise be faced with a test that is too difficult. The belief is that the alternate assessment increases the validity of the test scores and the contention that test levels should be matched to the student's IEP goals and objectives (Thurlow, Elliott, & Ysseldyke, 1999). Opponents of alternate assessments argue that the purpose of standardized assessments is to have high expectations for all students and to hold all students to the same standard (Thurlow et al, 1999). There are concerns about the valid interpretation of alternate assessments scores and problems with aggregation of alternate test scores for group reporting purposes. According to Thurlow and Minnema (2001), these issues have not been resolved through empirical research with no definitive data supporting the use or non-use of alternate assessments with students with disabilities.

Ideally, an alternative assessment requires students to actively accomplish complex and significant tasks, while bringing to bear prior knowledge, recent learning, and relevant skills to solve realistic or authentic problems rather than requiring students to choose from a set of provided responses. Exhibitions, investigations, demonstrations, written or oral responses, videotapes, journals, and portfolios are examples of the assessment alternatives we think of when we use the term "alternative assessment."

Research by Quenemon et al (2002) and Thompson & Thurlow (2003) indicates most states commonly use four approaches to alternate assessments. These include portfolio assessment, performance assessment, IEP linked body of evidence, and traditional test formats. States developed assessment criteria based on four assumptions that are the foundation of alternate assessments. These assessments should:

1. focus on authentic skills and on assessing experiences in community and other real life environments;
2. measure integrated skills across domains;
3. use continuous documentation methods, if at all possible; and
4. include as critical criteria the extent to which the system provides the needed supports and adaptations, and trains the student to use them

(Ysseldyke & Olsen, 1997).

Thompson and Thurlow (2001) indicate the most common approach to alternate assessment is a portfolio. Student portfolios are a systematic collection of student work that is evaluated and measured against a predetermined scoring criteria. Development of a portfolio assessment for students with the most severe disabilities requires the same thoughtful application analogous to what occurs for general assessment development (Quenemoen, Rigney, & Thurlow, 2002). Portfolios are typically designed so that data could be collected on the educational progress and accomplishments of students with the most severe disabilities. For assessment purposes a portfolio is collection of work, usually drawn from students' classroom work. A portfolio becomes a portfolio assessment when (1) the assessment purpose is defined; (2) criteria or methods are made clear for determining *what* is put into the portfolio, *by whom*, and *when*; and (3) criteria for assessing either the collection or individual pieces of work are identified and used to make judgments about performance. Portfolios can be designed to assess student progress, effort, and/or achievement, and encourage students to reflect on their learning. The results should be designed for use during the school improvement planning process,

to promote access to the general curriculum and to achieve proficiency levels of student performance based on state content standards (Thompson et al., 2001). A portfolio assessment allows students with the most severe disabilities to work toward the same content standards as all students using alternate learning expectations to measure progress.

States have indicated a change in the alternate assessment process since 2001. In 2001, 48% of the states used portfolio assessments as compared to 46% in 2003. The use of checklists or rating scales has increased from 18% in 2001 to 30% in 2003. Other assessments used include IEP analysis: 6% in 2001 and 8% in 2003; local selection of an alternate assessment: 24% in 2001 and 3% in 2003; and assessment under development or revision: 4% in 2001 and 6% in 2003 (Education Week *Quality Counts 2004*).

Eighty percent of state alternate assessments are aligned with the state academic standards content either through grade level, expanded standards, or through a combination of the content areas and functional skills. Four percent of the states conduct alternate assessments that are not aligned to the state standards and 3% allowed IEP teams to select the content used in the alternate assessment process (Thompson & Thurlow, 2001).

The connection to achievement level descriptors between the alternate assessments and the general education assessment has improved since 2001. Sixty-two percent of the states reported using the same descriptors in 2003 as compared to 36% in 2001. New Jersey uses the same descriptors of advanced proficient, proficient, not proficient.

Assessment Validity and Reliability

Validity refers to the quality of information an assessment yields. It is not an attribute of a test, but an attribute of a specific inference or conclusion based on test scores. If a test is not valid for the purpose used, then it has no inherent value. Traditional measurement research includes three aspects of validity: content validity, criterion-related validity, and construct validity. Construct validity is defined as encompassing the value and social implications of assessment results and is frequently referred to as consequential validity. Content validity refers to how adequately a test samples knowledge and skills that assess the approved written version of the curriculum or content area standards. A test is said to have criterion-related validity if its results parallel some other external criteria. The test results should be similar to the results of other assessments or measures of the same domain.

Large-scale assessment validity and reliability goes through a rigorous review lasting multiple years prior to the implementation of the test. Test items reflect educator and measurement experts understanding of student outcomes and scoring of student responses. Testing companies and state personnel review the test's content specifications, defining the specific skills and knowledge that will be tested and how the students will demonstrate their ability to use the knowledge and skills. Decisions are made on the depth of content coverage and the complexity of the test items. Decisions are made concerning how the students' responses are scored and the scoring criteria. The number of correct and incorrect answers on the test items provides the information needed to identify which students have achieved the knowledge and skills designated by the assessment. Due to years of experience in developing large-scale assessments, the

validity and reliability is rarely questioned or discussed by the public. It is assumed the tests meet the Standards for Educational and Psychological Testing (1999) as developed jointly by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME).

Previous low expectations for students with severe disabilities have created gaps in knowledge and skills. Until the recent requirement for inclusion of students with disabilities in large-scale testing, educators, measurement experts or test companies had direct experience in defining what successful outcomes are for this group of students at least as these outcomes are defined through measurement of academic achievement (Quenemeon, Thompson, & Thurlow, 2003). Questions have been raised about how to develop assessments and implement the assessments that are appropriate for all students without weakening the content standards and minimizing the rigor of performance (Almond, Quenemoen, Olsen, Thurlow, 2000). Concerns have surfaced as to whether the assessments for these students can be used as an effective measure of student progress based on state content standards and the validity and reliability of these tests.

An assumption of standardized tests is that state assessments measure knowledge and skill within content area in the same way for both students with disabilities and nondisabled students. This assumption has yet to be explored through empirical research (McDonnell et al. (1997). Before making assumptions about the achievement of content standards by students with disabilities based on statewide assessments results, examination to determine if those assessments are valid and reliable for these students is

necessary. Since alternate assessments are a more recent development than general assessments, the assumptions underlying them continue to be discussed and researched.

Most states currently do not have data on technical quality of the assessments due to the newness of the alternate assessment process. Researchers have questioned the reliability and validity of alternate assessments. These studies suggest that the concepts of reliability and validity do not have precisely the same meanings for portfolio assessments as they do for pencil-and-paper assessments. These studies explored the interaction of assessment and instruction and call for rethinking the quantitative models of reliability and validity and work toward a more qualitative method of assessment that establishes reliability and validity. Koretz, Stecher, Klein & McCaffrey (1994), suggest the difficulty in establishing reliability and validity of an alternate assessment lies in the values of the state in defining student success. Koretz, et al. (1994) state a conflict that arises between the instructional and measurement goals that either embed assessment in the instructional process or rely on nonstandardized tasks affects the assessments reliability and validity.

Perhaps the lack of evidence of the technical quality of alternate assessment has resulted in the states hesitancy to include the test results in the accountability formulas. Since large-scale assessment is a relatively new advancement in special education, studies and recommendations to assess reliability and validity in large-scale assessment in special education are just now occurring (Tindal et al., 2003).

Recent research on alternate assessments has focused on examining teacher opinions about the portfolio (Kleinhert, Kennedy, & Kearns, 1999), correlating teacher opinion and student scores (Kampfer et al., 2001), and examining the impact of the IEP

development (Kleinert & Kearns, 2001; Kleinhert, Kennedy & Kearns, 1997). Stiggins (2000) suggests that professional and subjective judgment is key to determining the quality of portfolio assessment. He proposes that a teacher's vision of what it means to be academically successful affects the results of the performance assessment. The teacher's understanding of the assessment and the student achievement is key to improved student achievement. Although this research has been somewhat beneficial when investigating the impact of portfolio assessments on instruction, it does not address the issue regarding the reliability and validity qualities of alternate assessment for use as part of the states accountability system.

Studies have investigated if the amount of time a teacher puts into preparing the alternate assessment, specifically a portfolio, affects the student's score. In 1999, Kampfer, et al. (as cited in Kleinert and Kearns, (2001) p. 221) surveyed all the special education teachers in Kentucky that had students participate in the alternate assessment. They investigated (a) the relationship between the amount of time spent working on the portfolios and the score; (b) which portfolio items required the most teacher effort; (c) to what extent teacher variables such as experience and training and instructional variables such as the amount of student involvement in the construction of the portfolio, portfolio elements embedded in instruction and the perceived benefit to the student predict the portfolio score; (d) and what aspects of the alternate assessment tend to be of most concern for teachers. The results indicate that instructional variables, teacher variables, and time spent on preparation accounted for 27.5% of the variance in scores. However, the instructional variables were the critical variables in predicting the variance in the scores accounting for 24.1% of the variance, not the amount of time spent on preparing

the portfolio or the teacher's experience. These results may indicate that teachers need more concrete strategies for embedding portfolio development into daily instruction and strategies to increase the student's involvement in the creation of the portfolio.

By the end of the 20th century, many states had developed alternate assessments as required by IDEA' 97 although many states were uncertain as to how they would incorporate the results of alternate assessments into accountability formulas (Quenemoen, et al., 2002). The passage of NCLB 2001 mandated states to develop a process so the results of the alternate assessments can be aggregated with those of the state general assessment. However, before those performance standards can be set, test developers need to assure that the alternate assessment is a valid measure based on the state concept. Messick (1989, as cited in Linn) outlined five types of evidence for determining validity: content, response process, internal structure, relations to other variables, and consequences. Miller and Linn (2000) discussed how Messick's framework applies to performance assessment and concerns with performance assessments such as the limited number of items assessed, the alignment of the assessment with the content standards, the confounding of constructs across content areas, and the influence of the scoring method on the construct and how these concerns should be addressed to collect validity evidence.

Content validity by itself is not considered sufficient evidence for assessing the validity of a test; however, the content evidence is important because the content of each portfolio varies based on the student's instructional program and the teacher's decision on what to include. Miller and Linn (2000) state that performance assessments present challenges with regard to content evidence because assessments are not always aligned with the content standards, and there are a limited number of tasks assessed which

narrows the definition of the construct. In a study of the Washington Alternate Assessment System, Johnson and Arnold (2004) found that the lack of construct validity in a portfolio assessment system may be due to teachers being unclear about how to select IEP skills that are connected to the state content standard areas. The student's curriculum may not be aligned with the state content standards because the state standards do not represent functional goals for the student. Each portfolio also contained a limited number of skills assessed to represent a content area. Therefore it raises the question if this limited number of skills assessed sufficiently represents the broad content area.

Educators and policymakers have questioned whether the alternate assessments are assessing the depth of the concept behind the content standard or simply the performance of a skill. Concerns focus on whether scores of an alternate assessment based on work place skills measure the same progress toward core academic content standards and the ability for these scores to be aggregated within a state's scores (Almond, et al., 2000). According to the Education Week *Quality Counts 2004*, thirty-two states assess students with severe disabilities on extended or expanded content standards, six use grade level content standards, two use functional skills not aligned to the content standards, and four use a combination of functional skills and content standards.

According to the American Educational Research Association (AERA) (2000), the response process of test takers can provide evidence concerning the connection between the construct and the performance of the test taker. In a portfolio assessment for students with severe disabilities, the teacher is responsible for assembling the data, therefore the student's score is based mainly on the teacher's presentation of the data and the

completion of the process. The teacher responds to the state directive to select skills based on IEP objectives, creating the environment and supports necessary for the student to practice and measure student progress, collects the data, and presents the data in a format designated by the state. The scores in a portfolio are open to interpretation including the appropriateness of the student's IEP goals, the effectiveness of the instructional program, the alignment of the core content standards to the curriculum, or the need to provide effective staff development for teachers on content based instruction (Johnson & Arnold, 2004). As a result it could be reasoned that the teacher's response process will have an enormous impact on the student's score.

The internal structure of a test can indicate the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based (AERA 2000, p.13). Therefore, the selection of a scoring method is key to understanding the inferences that can be made from the portfolio assessment (Miller & Linn, 2000). Alternate assessments are typically scored through use of a rubric. States use a similar approach to setting up a rubric in that there are common criteria used to measure quality and the level of success of the student portfolio. Thompson & Thurlow (2001) suggests the criteria should measure how well a student can perform target behaviors or skills; the level of independence a student demonstrates; how well the skills are generalized to different environments, with different people, or in different activities; and how appropriate the skills and activities are for a student (p.104). Skill generalization is an important assessment criteria since it provided evidence the student is working toward achieving at the same level of the non disabled peers and is meeting the core content standards.

Kleinert and Kearns (2001) studied the importance of scoring, specifically the standards or performance criteria used, in the development of alternate assessments. More than 20% of the 44 national authorities surveyed stated that scoring criteria must not be scored in isolation but should be used concurrently to be meaningful (Kleinert & Kearns, 2001, p.215). Targeted skills should be evidenced across multiple settings and community settings, and age appropriate choices should occur in the context of activities with peers without disabilities are examples of cross scoring criteria. Alternate assessment scoring varies and reflects teacher generated checklists in which the teacher scores their own students and submit the scores to the state with no state oversight which may or may not be followed by a sample percentage of student work samples. Other states train teachers to score the student work, but not a teacher's own students or own district work, followed by a scoring of sample work. In other states, a regional panel reviews all portfolios with at least two independent scorers.

Scoring of portfolios is, according to Salvia and Ysseldyke (2001), "neither simple or straightforward" (p. 248) nor is score aggregation or score reliability issues resolved. Salvia and Ysseldyke suggest that attention to greater objectivity, less complexity, more scorer training, and greater comparability of portfolio contents are the keys to better practice (p. 257). Thompson & Thurlow (2001) suggested states need to be aware of the following challenges that may affect the validity and reliability of the assessments. Portfolio assessments may have lower reliability and comparability of results than other types of tests. They are more difficult to implement and present greater variability such as time spent on the development of the portfolio, training of teachers, and teacher

support, and they are more difficult to score, more time consuming to score, and more subjective to score than standardized tests.

Creating portfolio assessments that are reliable and valid is costly and challenging when used to measure progress for students with disabilities. Measurement error is consistently greater than that associated with traditional tests (Koretz, et al., 1994). Portfolios typically integrate a variety of knowledge and skills to produce a product. The interdisciplinary characteristics of portfolio assessment tasks increases the probability that the severity of the disability will interfere with the validity of inferences made. For example, for students with processing to cognitive impairments it may be difficult to disentangle a reading skill from the vocabulary that is part of the standard, leading to the question of if the test assesses the determined outcome.

To assure greater validity and reliability, portfolios should contain pre-defined structure for the portfolios and require a number of entries and types of entries to assure all of the portfolios are scored using the same criteria (Olsen, Mead, & Payne, 2002; Quenemon, Rigley & Thurlow, 2002). Scoring criteria can be a direct measure of student achievement, a measure of system conditions necessary for student success, or a combination of the two. Unlike traditional large-scale tests with scoring criterion embedded in the test development process, portfolios tend to represent a continuum of scoring approaches as opposed to specific categories such as appropriateness for grade level and degree of difficulty. The scoring criteria should reflect the state's identified outcomes as evidenced through sample student work or in student response items on a test. Scoring criteria are often in the form of rubrics and include: variety in settings, student progress, ability to generalize, age appropriateness, alignment to the standards,

amount of support, self determination, socialization opportunities, general education participation, and parent satisfaction (Thompson & Thurlow, 2001).

For the purposes of result comparison, Quenemoen, et. al. (2002) suggests six commonly used scoring criteria regardless of the assessment protocol used by the states. These common areas include: independence, generalization of skills, appropriateness of skills, performance, IEP linkage, and content standards linkage. Quenemoen, et.al (2002) compared and contrasted the assumptions and values embedded in scoring criteria used in five states for their alternate assumptions. Although these five states use different alternate assessment approaches, they share three basic quality indicators. Each state developed their alternate assessment through input from the stakeholders including educators, parents, researchers, and technical advisors reflecting professional and research based understanding of possible outcomes for students with significant disabilities. Each state continues to improve the reliability and validity of their assessment through documentation of the technical adequacy of the assessment. The states were also able to document a coherent alignment of the basic assumptions about teaching and learner outcomes. The methods used by the states reflect an understanding of best practices in alternate assessment leading to a level of confidence in the internal consistency and integrity by the researchers, thus providing a guide for other states' development of the alternate assessment.

New Jersey Assessment and Accountability System

New Jersey has used statewide assessments as a means for student progress and program planning for many years, even before the federal legislation mandate. In 1975, the New Jersey legislature passed the Public School Education Act (PSEA) “to provide all children of New Jersey, regardless of socioeconomic status or geographic location, the educational opportunity which will prepare them to function politically, economically and socially in a democratic society” (N.J.D.O.E.). In 1976, PSEA was amended to establish uniform standards of minimum achievement in basic communication and computational skills. This amendment provided the legal basis for use of a test as a graduation requirement.

From 1978 through 1982, third, sixth and ninth graders participated in the Minimum Basic Skills (MBS) testing to assess reading and mathematics skills. Beginning in 1981, the MBS became one of the requirements for ninth grades to pass in order to graduate. In 1983, New Jersey adopted the Grade 9 High School Proficiency Test (HSPT9) to assess reading, writing and mathematics. The HSPT9 was administered as a graduation requirement during the 1985-86 school year.

In 1988, the New Jersey Legislature replaced the HSPT9 with the High School Proficiency Test (HSPT11) and added the Grade 8 Early Warning Test (EWT). The EWT was to serve as an assessment to assist in student placement and program planning, while the HSPT11 was a graduation requirement for all students.

The Core Curriculum Content Standards (CCCS) adopted by the New Jersey Department of Education in 1996 and revised in 2003-2004, identify what New Jersey students should know and be able to do by specific grades. The New Jersey’s

Assessment and Accountability system is multi-level series of test developed to determine how a student is progressing toward achieving the CCCS. New Jersey's statewide assessment system has been aligned with NCLB 2001. The statewide assessment system is referred to as the New Jersey Assessment of Knowledge and Skills (NJASK 3 and 4), the Grade Eight Proficiency Assessment (GEPA), the High School Proficiency Assessment (HSAP), the Special Review Assessment (SRA) and the New Jersey Alternate Proficiency Assessment ("NJAPA"). It mandates that all students in Grades 3, 4, 8, 11 take the relevant test for the designated grade level in the spring of the year and earn a proficient status on the H.S.P.A to receive a standard high school diploma. Students unable to participate in the NJASK, GEPA, HSAP, are required to participate in the NJAPA (N.J.D.O.E.)

The statewide assessment is designed to inform the New Jersey Department of Education and the public about how well schools and school districts are performing in teaching the CCCS. The test informs districts and parents about a student's proficiency in the tested subject area. The primary purpose of the New Jersey assessment program is to assess student achievement in reading, math, and science. A secondary purpose of the New Jersey Assessment is to compare the performance of students in New Jersey to student's performance across the country and within school districts in New Jersey by similar district factor grouping (DFG) status. Additional components of the plan include measuring student's performance, reporting school performance, and rewards for or sanctions for poor performing schools.

In accordance with NCLB 2001, New Jersey students with disabilities must participate in each subject area assessment of the NJASK, GEPA, and HSAP unless the student's

IEP team determines that the student has not been instructed in any of the knowledge and skills tests and would not be able to do the types of items on the assessment (N.J.D.O.E.). The IEP team must decide what accommodations and/or modifications are necessary for the student to participate in the assessment and whether the student will take the general assessment or the NJAPA.

Students deemed unable to participate in the general assessment even with accommodations and/or modifications, participate in the NJAPA. The NJAPA is a portfolio assessment designed to measure progress for these students. Portfolio contents are individualized, linked to the CCCS, and may include a wide range of samples of student learning, including but not limited to actual student work, observations recorded by multiple persons on multiple occasions, test results, record reviews, or even video and audio records of student performance. The portfolio contents are scored according to predefined scoring criteria, usually through application of a scoring rubric to the varying samples of work.

In New Jersey, the majority of students with disabilities participate in general assessments as opposed to alternate assessments. In 2004, a total of 3,314 students took the APA with 3,039 students having valid Language Arts Literacy portfolios and 2,978 having valid Mathematics portfolios. New Jersey state results for the APA indicate that the percentage of students that scored at or above proficient on the Language Arts Literacy portfolios was: Grade 3, 91.7%; Grade 4, 92.9%; Grade 8, 91.3%; and Grade 11, 90.5%. Non-special needs districts scores and average of 93.5 % at or above proficient on the Language Arts Literacy portfolios for Grades 3, 4, and 8 and 90.7% in Grade 11, while special needs district students scored an average of 87.6% in Grade 3, 91.4% in

Grade 4, 86.6% in Grade 8 and 89.6% in Grade 11. Mathematics APA portfolio results indicate 87.9% of students assessed at Grade 3 were at or above proficient with 4th Grade results at 88.5%, 8th Grades results at 89.4% and 11th Grade at 86.7%. Non-special needs districts scores and average of 90 % at or above proficient on the Mathematics portfolios for Grades 3, 4, and 8 and 85.7% in Grade 11 while special needs district students scored an average of 83.4% in Grade 3, 86% in Grade 4 and Grade 8 and 89.7% in Grade 11. Since this is the first year that scores were formally reported for public distribution caution should be used when concluding the success of the results of the testing program (N.J.D.O.E.).

Effects of Inclusion in Testing

Over the past 30 years, major educational changes have occurred in the education of students with disabilities. Education for all Handicap Act (EHA) did not envision one placement for all students, but a full range or continuum of alternative placements for students with disabilities. P.L.94-142 defines least restrictive environment as the following:

...to the maximum extent appropriate, handicapped children, including those children in public and private institutions or other care facilities, are educated with children who are not handicapped, and that special classes, separate schooling, or other removal of handicapped children from the regular educational environment occurs only when the nature of the severity of the handicap is such that the education in regular classes with the use of supplementary aids and services cannot be achieved satisfactorily (P.L.: 94-142, section 1412[5][b])

Least restrictive environment is clearly defined in the law; however, interpretations of least restrictive environment vary considerably. This differs from the idea of inclusion

and mainstreaming. Least restrictive environment is the legal term whereas the term inclusion does not appear in the law. The term mainstreaming and integration, relating to student placement, also do not appear in federal laws or regulations. Even though the law has defined least restrictive environment, educators have created definitions for mainstreaming, integration and inclusion.

The educational initiative for least restrictive environment has raised the awareness of the need to educate students with disabilities within the general education setting. More students with disabilities are now being educated in general education settings with their peers without disabilities. Although state departments of education, and legislation have endorsed the educational movement toward least restrictive environment, recent court decisions reinforce the IDEA that “one size does not fit all.” Schools vary in their placement of students with disabilities ranging from self-contained programs in separate schools to inclusion of student in general education classes. Ideally, students with disabilities should have full access to the general education curriculum and opportunities for interactions with typical peers. Theoretically, this should occur with the support of general educators and school administrators. However, without legislative mandate this was not always a reality (Kleinhert & Kearns, 2001).

Several important court cases have provided guidelines for including students. Daniel RR v. State Board of Education, 1989, Greer v. Rome City School District in Georgia, 1991, Sacramento City Unified School District v. Holland, 1992, and Oberti v. Board of Education of Borough of Clementon School District, 1993. The Oberti case provides the most relevant guidelines in New Jersey. The case allowed court to rule the following:

School districts have an obligation to consider placing students with disabilities in regular education classes with supplementary aids and services before they explore other alternatives... Therefore the starting place for an inquiry into what the appropriate least restrictive environment for a student must be inclusion as a regular member of a regular education classroom; then the inquiry can then move to the matter of a more restrictive environment if it is found that that environment is not appropriate.

Simply assuring a student is placed in a least restrictive environment does not always guarantee the student has access to the general curriculum and therefore access to quality instruction and standardized assessment. The phrase access to the general curriculum was defined in IDEA '97 referring to the need for the education of students with disabilities to be based within the general education curriculum and standards.

To assure the inclusion of students in the general education curriculum, IDEA'97 stipulates that the planning and development of the IEP must reflect the student's participation in the general curriculum. The IEP must include the student's present levels of education performance and how the student's disability affects the student's participation and progress in the general education curriculum (614[d] [1] [A] [I] [aa]). IDEA 2004 further stipulates that for students with disabilities who take alternate assessments aligned to alternate achievement standards, a description of benchmarks or short-term objectives must be included (614[d] [1] [A] [I] [cc]). A statement of measurable annual goals, including academic and functional goals designed to meet the student's needs that result from the student's disability to enable the student to be involved and make progress in the general education curriculum (614[d] [1] [A] [II] [aa]). The special education, related services, and supplementary aids must be designed to enable the student to be involved and make progress in the general curriculum in accordance with subclause (I) and to participate in extracurricular activities and other

nonacademic activities (614[d] [1] [A] [IV] [bb]). Through the present levels of educational performance, the goals and objectives and the provided services, a student with a disability is therefore assured involvement in the general education curriculum.

NCLB 2001 and the revisions to IDEA 2004 provide the legal mandates for the participation of students with disabilities in state and district assessments; however, the questions of how to accomplish this task and the benefits of this inclusion remains uncertain. Kleinert and Kearns (2001) suggest at least three strategies for ensuring the inclusion of all students with disabilities in the general curriculum. The first strategy is to ensure that the student has the opportunity to learn their IEP objectives in the context of the general education classroom and to receive instruction in the context of the general education activities (p. 7). The student should have opportunity to participate in these activities in the home, the school and within the community.

Assuring a student participates in activities in the general education classroom does not assure a connection to the learning standards. Owen White's 1980 (as cited in Kleinert and Kearns, 2001, p.7) approach called for alignment of the IEP objectives with the learning standards through critical function. When assessing students with disabilities, the focus is on the function of the standard that will enhance the students' life. This approach allows the student to meet the learning standards by demonstrating acquisition of the skills, thus mastering the essential content of core learning for all students (Kleinert & Kearns, 2001, p.7).

The third strategy to enable students with disabilities to access the general curriculum is to address the standards through learning the academic content. Through an inclusive

curriculum design, students with disabilities would be a part of the general education classroom to learn at least some of the academic content (Kleinert & Kearns, 2001, p.8).

When the inclusion of students in state assessment systems was mandated, it was viewed as a process to assure students with disabilities would be included in local and state policy making decisions. The inclusion was viewed as a method to improve opinions about students with disabilities, improve access to the general education curriculum, and improve instruction in special education programs (Browder, Spooner, Algozzine, Ahlgrim-Delzell, & Karvonen, 2003). Research is beginning to show positive consequences of this inclusion although extensive studies on the effects of inclusion in statewide testing and the accountability process is limited due to the recent legislative requirements.

Data on the influence of alternate assessment results on policy decisions is difficult to determine at this time due to schools and states not having a clear picture of educational outcomes related to the inclusion of students in the assessment process. In a report on state reporting processes in 2001, Thompson and Thurlow (2001) reported that only half the states reported alternate assessment scores with the remaining states reporting that they were continuing to decide how to report the scores. The reporting of results has improved as reflected in Thompson and Thurlow's (2001) state report that indicates 96% of states reported student scores. At this point the influence of inclusion on policy and decision making is inconclusive until additional data and information is obtained.

Thompson and Thurlow's (2001) survey of state directors of special education found positive effects among states with an increase in the inclusion of these students in state assessments (90%). Questions surrounding the inclusion of students in the general

assessment or the alternate assessment have generated research in determining eligibility requirements for participation in alternate assessments. During the initial development of alternate assessments, 9 out of 12 states considered using cognitive level and adaptive behavior as precluding students from participation in the general assessment. Bechard's (2001) report on models for reporting results of assessments found that 20% of students with disabilities participated in the alternate assessment instead of the general assessment. Kleinert and Kearns (as cited in Browder, Spooner, Algozzine, Ahlgrim-Dezell, & Karvonen, 2003), reported that in Kentucky the alternate assessment rate of participation was much lower and used specifically for students with the most significant disabilities.

Although the inclusion rate has improved, questions remain as to whether participation has increased student expectations. Indirect evidence is reported in Kentucky by Kleinert and Kearns (1999, as cited in Browder, Spooner, et al., 2003). Scores on the alternate assessment have improved over the years with teachers reporting that the expectation for students to complete self-determination skills as part of the alternate assessment process has improved the expectation that these students are capable of self-assessing. The study reports an increase in the use of augmentative communication systems alluding to the expectation of the need for the students to communicate and express themselves.

An option for determining if alternate assessments have affected expectations for students is to review student IEPs. Thompson and Thurlow (2001), found in 41 states, only 8 states had no reference to the alternate assessment in the IEP and only 5 states addressed the state standards. Turner, et al. (2000) found no correlation between the

quality of the IEP and performance on the alternate assessment or a reflection of expectations on the IEP.

Relationship of the alternate assessment to the state's standards is another technique used to determine student expectations. Researchers found including academics rather than just life skills or functional skills in alternate assessments linked the state standards to the alternate assessment process (Kleinert & Kearns, 2001; Thompson & Thurlow, 2001).

The premise that the use of alternate assessments will increase access to the general curriculum and assess all students on the state standards is not evident in the research. Browder, Spooner et al., (2003) found only 17% of states evaluate whether the skills included in the alternate assessment reflect state standards.

Data received from the alternate assessment to improve the quality of instruction is necessary to determine the benefits of the alternate assessment process. Thompson and Quenemeon et al. (2003) state the time invested in the alternate assessment may benefit both the student and the teacher. The information provided through the assessment results can be used to improve instruction thus improving student achievement.

There is currently insufficient research to state specific benefits of the alternate assessment on improved instruction. One study by Turner et. al (2000) correlated time investment to quality indicators and ultimately improved instruction and improved student outcomes. Others question if the amount of time spent on the preparation and collection of materials for the portfolio stresses teachers without necessarily improving instruction. DeStefano et. al. (2001) conducted a study the benefits of staff development focusing on the planning and participation in different assessments. The study showed

improvement in the teachers' understanding of when to use the alternate assessment but not on how to use the results to improve instruction. Browder & Spooner, et al. (2003) described two methods to use the data for decisions regarding instruction. They stated that unless teachers are specifically trained on data based decision making, the benefits to improved instructional process through use of the alternate assessment may not occur.

States are beginning to be able to document performance trends of students with disabilities. Forty of the fifty states (80%) report the ability to document performance trends (Thompson & Thurlow, 2001). The reported trends show overall improved state assessment performance as compared to the trend data from 1998 to 2001 that showed progress in reading and writing but minimal progress in math. Trends have been difficult to compare to past years of assessment progress due to changes in the assessment process over the past several years.

Though the progress of students with disabilities in standards-based accountability has demonstrated improvement, states continue to face challenges in an effort to increase student achievement and administer assessments that provide documentation of this achievement. The major concern among 48% of the state directors of special education is that students with disabilities will not be able to achieve proficiency on state assessment. Twenty-six percent of the state directors believe the students cannot access or reach the states standards, while 28% believe that students with disabilities make school look less effective due to the low test scores. Other concerns include students being stressed by taking tests (40%), students with disabilities not graduating (20%), and to a lesser extent

an increase in students identified for alternate assessments (12%) (Thompson & Thurlow, 2001).

Summary

Students with disabilities are required by NCLB 2001 to participate and demonstrate improved performance in state assessments. The assessments are used to provide information on the educational progress of students and the extent to which students are achieving state standards. Since the start of alternate assessments, students eligible for these assessments ranged from students with severe and profound disabilities to some students with moderate and other disabilities. Gathering data on the performance of students through alternate assessments requires rethinking traditional assessment methods. Reporting on students with disabilities must be done in the same way and with the same frequency as for students without disabilities. IDEA 2004 reinforced these requirements, clarifying that states must report on both alternate assessments based on grade-level achievement standards and alternate achievement standards, and must also report on the number of students using accommodations to participate in the general assessment. States are now implementing accountability plans with consequences assigned to schools, administrators, teachers, and other educators. A review of the current research in large-scale assessment and accountability revealed that studies to date have not been definitive on the consequences of the assessment and accountability reform effort on students with disabilities with respect to the curriculum, instruction, parental and teacher awareness.

Chapter III

Methodology

Overview

This chapter describes the methodology of the study including the research design, participants, survey and focus group process, data collection procedures, and the analysis of the data. Qualitative and quantitative research were chosen to analyze the insights, attitudes and experiences of directors, supervisors, or coordinators of special education programs in public school districts in New Jersey toward the consequences of large-scale assessment results for students with disabilities. The results provide an understanding of the initiatives, trends, accomplishments, and emerging issues involving the New Jersey alternate assessment process. A second purpose was to determine if the accountability requirement has had an effect on curriculum, instruction, and special education placement.

A review of the current research in large-scale assessment and accountability reveals that studies to date have not been definitive on the consequences of the assessment and accountability reform effort on students with disabilities with respect to the curriculum, instruction, parental and teacher awareness. This study explored directors of special services' perceptions regarding the consequences of high-stakes assessment for students with disabilities, with specific reference to the New Jersey Alternate Proficiency Assessment (NJAPA) and the impact of accountability on special education services.

The NJAPA assesses the achievement levels of New Jersey's students with the most severe and cognitive disabilities and is the basis for New Jersey's accountability

system relative to these students. The primary intent of the New Jersey assessment system is to improve student achievement levels and increase accountability for school districts in assuring adequate yearly progress and quality instruction. The outcomes of the assessment policy are considered indicators of how New Jersey's assessment program is achieving its intended purpose, as well as how it is functioning beyond its intended purpose of student achievement. Evidence of improved student achievement is an indicator used to determine a school's ranking and the School Report Card and as a performance based assessment for funding.

Although test scores are indicators of student achievement, they cannot alone measure the impact of the New Jersey assessment process. The literature indicates large-scale assessments and accountability systems impact a variety of outcomes including student learning, instruction, curriculum development, student, teacher and parental awareness, and professional development for instructional and administrative staff. Most research reflects national trends or is relevant to specific states. Therefore, research conducted within the central and northern New Jersey region has been conducted to identify consequences that are unique to this area as well as trends that are reflective of other state and national movements.

Methodology

The descriptive study employed two methodologies: administering a survey (quantitative design) and conducting a focus group session (qualitative design), to explore directors of special services' perceptions regarding the consequences of high-stakes

assessment for students with disabilities, with specific reference to the NJAPA and the impact of accountability on special education services.

Research Design

Evidence to evaluate large-scale assessment systems has been identified by the American Education Research Association (2000), The National Center on Educational Outcomes (Thompson & Thurlow, 2001) as well as individual research (Chudowsky & Behuniak, 1998; Lane, Parke, & Stone, 1998; Meherns, 1998). Lane et al. (1998) suggests the first step in examining the consequences of an assessment system is to identify the intended effects on such variables as the curriculum, the instructional contents and strategies, and content and format of classroom assignments, and so on. Examining the intended positive effects will likely lead to the identification of negative effects and unintended consequences.

According to Fink (1995, p. 1), the survey is a direct research method used for collecting information to describe, compare or explain knowledge, attitudes, and behavior. Rea and Parker (1997) state that “there is no better method for determining, with a known level of accuracy, information about large populations” (p.3). A survey should be designed to meet the objectives of the research obtain the most complete and accurate information possible; and to be completed within the limits of available time and resources (Sheatsley, as cited in Rossi, Wright, Anderson, 1983). Based on Fraenkel’s (1996) and Linsky, (1975) submission that a survey instrument should be brief enough to encourage the subjects to complete and return it, the researcher attempted to word the survey items clearly and briefly so that the participants could easily understand the

question. Assuring the results obtained from the sample population can be generalized to the “target population” is extremely important when defining the population of the study (Anderson, 1999, p.196). The researcher must also select the sample population that has information the researcher is seeking and is willing to respond to the survey (Fraenkel, 1996).

The Perceptions of Accountability in Special Education (PASE) is a Likert type scale consisting of 15 questions. Questions focused on student instruction, testing accommodations, alternate assessments, accountability, and current and emerging curricular practices and issues. Each respondent was asked the same set of questions to ensure reliability.

The choice of the demographics items was included in the survey since peoples’ perceptions are affected by demographic factors such as their sex, age, level of education, socioeconomic factors, ethnicity, and contact (Sheatsley, as cited in Rossi, Wright, Anderson, 1983, p. 204). Demographic questions in this study’s survey include personal information, such as years of experience as a director and gender. District demographic questions deal with the district size, total number of special education students as of the official 2004 count in New Jersey and the District Factor Group.

The District Factor Group is an indicator of the socioeconomic status of a district’s population. The measure was developed in 1974, based on the 1970 census. Revisions occurred in 1984 and 1992 as a result of the changes in the 1980 and 1990 census. The variables considered by New Jersey include: percent of adult residents who failed to complete high school, percent of adults who attend college, occupational status of household members, population density, income, unemployment and poverty. After

applying a formula known as the principal components analysis, a single socioeconomic measure is obtained for each district. The districts are broken down into eight categories, A, B, CD, DE, FG, GH, I, J, with A the lowest socioeconomic district and J the highest (N.J.D.O.E.). For the purpose of this study, the groups were ranked low (A, B), middle (CD, DE, FG, and GH), and (I, J) high.

Issues considered in the survey design included the clarity of the directions and questions, the length of the survey and the order of the items. According to Babbie (1999), the order impacts the less educated more than the more educated. Since the population surveyed reflects a highly educated population, Babbie's suggestion is not paramount to the design of the survey. The answer to a question is valuable to the extent that it can show a predictable relationship to facts or subjective areas of interest. A survey meets its objectives when the relationship between the answers and the purpose of the study is evident (Fowler, 1993). Selection of the sample population can also affect the results of a survey even if it appears the selected questionnaire is the most suitable for the research purpose. In a survey, it is assumed that the respondents have a general knowledge of the subject. However, Sheatsley's (as cited in Rossi et al., 1983) suggestions that if the population does not have the information, they cannot recall it, cannot predict it, or are ignorant in the subject matter, the end result may not be applicable to the research purpose (p. 196). Since this research directly implicates a subset of a larger population, the concerns related to a lack of knowledge of the subject area has been controlled (Fink, 1995, p.27, Fraenkel, 1996). The sample group has a clearly defined eligibility in that the participant must be a director, supervisor, or coordinator of a special education program in a public school district in New Jersey with

an administrative and functional knowledge of special education services. The willingness or readiness of the respondent to reply may also impact the findings. Careful introduction of the survey, proper survey auspices, and a well-planned line of questioning that does not depend on a few blunt questions add to the validity of the questionnaire (Sheatsley, as cited in Rossi et al., 1983).

Fink (1995) describes validity of an instrument as being accurate, while reliability of an instrument is described as being consistent. The relationship of the intent of the questions to the study's purpose, and the consistency of the responses across the reviewed surveys were followed to assure a relationship to the purpose of the study.

Focus Group

In addition to the survey, the researcher conducted a focus group to add to the richness of the data. Because the subject of accountability and assessment are both sensitive and complex, the choice of a survey and focus group was chosen. Survey researchers frequently are concerned that the use of one research method such as a survey limits the study results and a less acceptable study. Campbell as cited in Rossi, et al., (1983) advocates for the use of multiple methods in the measurement of social variable. The use of triangulation or the method of combining methodologies is in the words of Patton, (1990) "Ideal" (p.187). Patton states "studies that use only one method are more vulnerable to errors linked to the particular method (e.g., loaded interview questions, biased or untrue responses) than studies that use multiple methods in which different types of data provide cross-data validity checks" (p. 188).

Chudowsky and Behuniak (1998) recommended focus group methodology in addition to survey methods for collecting evidence regarding consequences of testing systems in that focus groups (a) allow for unanticipated responses, (b) allow for the researcher to gain an understanding of the nature and strength of the emotional reactions to participants, (c) provide illustrative points and added interest to the reader, and (d) produce insights by the nature of the interaction of the group members.

Through focus groups, the researcher derives understanding based on the discussions, as opposed to the testing of a theory or hypothesis. Focus groups may be defined as group interviews that bring together a small number of participants from a well-defined target population to discuss a set of preselected topics under the guidance of a moderator (Krueger, 2000). Focus groups typically elicit a wide range and depth of responses and are often used to evaluate ongoing programs (Patton, 1990). Information obtained from the focus group can be maximized by proper selection of the focus group participants, planning of the group interview time, clarification of any issues, and systematic analysis of the information obtained during the discussion.

The questions for the focus group were determined and sequenced based on an analysis of the literature, focus group questions from previous doctoral dissertations, and based on open-ended survey questions from the National Council for Educational Outcome research. The question guide was arranged in a logical manner to facilitate group discussion with each type of question designed to provide feedback relative to the research questions. Questions were open-ended and non-specific so that the participant can interpret the questions based on his or her own experiences. Opening questions and introductory questions were designed to establish trust with the group members and

provide an understanding as to why they were an essential part of the group. Ending and summary questions provided a check to the interviewer the participants perceived responses. These questions also provided an opportunity to identify the most important aspect of their beliefs. Probes for clarification of vague and ambiguous responses, probes for elaboration on potentially important points, and silent probes were utilized (Murphy 1980).

The researcher's professional relationship was considered when developing the procedure for conducting the focus group. Professional contact with the directors of special services occurs occasionally during state and county special education meetings and during the placement of a student at the researcher's school. Professional contact with the directors of special services will not compromise the information obtained from the data. To reduce bias, participants were given a code that was referenced when recording of the observations. Participants were selected who represent diverse district factor groupings will reduce participant conformity. Providing a post survey so participants had the opportunity to reflect on their individual experiences in the event they were unable to respond during the discussion or to relay concerns they were unable to respond to during the group will reduce bias.

Subjects

PASE Survey Participants

According to the New Jersey Department of Education, there are 118 districts in Essex, Middlesex, Morris, Somerset, Sussex, Union, and Warren counties in central and

Northern New Jersey. Districts in central and northern New Jersey were selected based on the district's 2004-2005 student enrollment, the District Factor Group (DFG), the district's listing as a K-8 or K-12 school district, and the district's participation in the New Jersey Assessment of Skills and Knowledge for Grades three-four (NJASK 3-4), the Grade Eight Proficiency Assessment (GEPA), and the Alternate Proficiency Assessment (APA).

Of the original 118 districts, four participated in the pilot study and 14 districts of the 118 did not meet the determiner requirements and were eliminated as possible participants. Surveys were sent to 100 directors of special services representing each department of special services in the designated districts. The information about the districts was obtained from the New Jersey Department of Education website. If available, the surveys were directed to the specific director of special services. The distribution of the groupings in this study are as follows: 13 districts in DFG Group AB; 59 districts in DFG Group CD, DE, FG, GH; and 46 districts in DFG Group I, J.

Although the position of the director of special services for each district is an administrative one requiring at least a supervisory certificate, there are many titles for this position. The title is indicative of the director's background. A background in special education as a teacher, supervisor, child study team member, learning disabilities/teacher consultant (LDTC), or a principal with special education background comprised the participants' backgrounds. Director of special education, director of special services, director of pupil personnel services, director of personnel services, supervisor of special education, coordinator of special services, coordinator of child study team, and

coordinator of special education are the common titles for this position in New Jersey. Even though the title differs among districts, the responsibility of operations for the district's special education program was the same.

Focus Group Participants

Participants of the focus group were directors of special services who have direct contact with the members of IEP team, including parents, case managers, general education and special education teachers. The directors are also responsible for monitoring curriculum and instructional implementation and reporting and analyzing reports and scores. The focus group was comprised of participants from K-8 and K-12 districts. The group was homogeneous in nature with respect to certification, but heterogeneous with respect to years of experience in education and DFG (Krueger, 2000).

To standardize the selection process, participants were chosen from a list of directors of special services from central and northern New Jersey. Criteria for participation included the following:

1. a valid New Jersey certificate as a principal, supervisor, administrator;
2. direct responsibility for supervision of IEP teams;
3. direct responsibility for reporting and analyzing assessments scores and data;
4. direct responsibility for program changes and implementation; and
5. 3 or more years experience in a supervisory position preferable special education

Data Collection

PASE Survey

The PASE survey was sent with a 2-week return time. A follow-up mailing was sent out with a 1-week deadline. A cover letter, self-addressed stamped envelope, and survey was included in the mailing. A separate self-addressed post-card was included with the initial survey for the participant to return if they want a summary of the results of the survey. The follow-up mailing included a different cover letter, no postcard, another survey, and self-addressed envelope.

Respondent's names were not requested on the surveys. In order to maintain anonymity of participants, a code number was written on the back of each survey. Each survey was coded numerically for the sole purpose of identifying directors who did not respond so that a second survey could be sent. As completed surveys were received, respondents were deleted from the mailing list. Surveys were returned to a P.O. Box located at the local post office of the researcher's place of work. A master list and all responses were maintained in a locked file located in the researcher's office to protect the confidentiality of the survey respondents.

Response bias and nonresponses were minimized through the selection of the sample population (Fraenkel, 1996). Consideration was given to inducing respondents to complete the questionnaire without intervention of the researcher; however, an incentive was not included in the survey mailing.

Focus Group

The researcher contacted potential focus group participants by mail and e-mail 10 to 14 days before the scheduled focus group. A confirmation letter or e-mail was sent 1 week prior to the focus group to confirm participation.

The focus group was recorded through written notes (as explained in the Letter of Solicitation and Informed Consent) and transcribed by the researcher. A summary of the collected data appears in Table 18 in chapter 14. Participants had the opportunity to verify the summary statement posed at the end of the session and through a summary of the discussion mailed to the participants.

Instrument Design

PASE Survey Design

A number of strategies can be used to improve validity and reduce bias with survey and focus group methodology. Systematic analysis of data is essential, from sequencing questions, capturing and coding of data, to participant verification of data.

The original survey PASE of 20 items and seven demographic questions was sent to a jury of experts. Included was the cover letter explaining the purpose of the research and the role of the jury of experts. The jury of experts was composed of four directors of special services who have been in the field of special education for 10 plus years and hold a doctoral level degree.

Each participant was asked to rate the items in the following manner:

1. Questions that you believe will reveal a director's perception toward the

alternate assessment process and accountability were to be given a+ (plus).

2. Questions that you believe will not reveal a director's perception toward the alternate assessment process and accountability were to be given a - (minus).

3. Questions about which you are unsure were given a 0.

The results of this rating system were complied with a+ (plus) receiving 2 points, a 0 receiving 1 point, and a - (minus) receiving 0 points. Survey questions were chosen and modified as a result of the jury of experts comments.

In addition, two open-ended questions were asked. The first asked for suggestions for the survey and the second for any other comments.

From the original survey, 15 items were chosen and modified as result of the jury of experts' comments. A pilot survey was sent to 10 directors of special services in selected districts. This mailing included a cover letter explaining the purpose of the study and the role of the pilot group. Two of the directors were also members of the jury of experts. The responses assisted in developing the final PASE survey. The following questions were asked to the pilot group:

1. How long did it take you to complete the survey?
2. Were the directions clear?
3. Were there any survey items that were unclear?
4. Should anything else be included in the survey?
5. Was the layout of the survey clear?
6. Were there any other comments?

Focus Group Design

A third party reviewer reviewed questions for appropriateness prior to the focus group meeting. Introductory questions were open-ended and designed to allow the participants to reflect on their own experiences and understanding of the problem that was being discussed. These “key questions” (Krueger, 2000) were open-ended and required the researcher to utilize probing to elicit more in-depth information.

Data Analysis

Analysis of the PASE Survey

Quantitative analysis procedures included basic descriptive statistics: measures of central tendency (mean and median) and dispersion (standard deviation and range).

Based on the research questions, three null hypotheses were developed to guide the quantitative portion of the research study.

1. There is no significant difference in the director of special services perception of accountability among the district size groups.

2. There is no significant difference in the directors of special services perception of accountability among district socio-economic status groups (DFG).

3. There is no significant difference in the directors of special services perception of accountability among the experience level groups.

Each hypothesis was tested using analysis of variance (ANOVA).

Analysis of the Focus Groups

Qualitative analysis included the coding of directors of special services comments according to themes and patterns. The researcher's professional judgment was used in coding based on an analysis of words used in the focus groups (word repetitions, key-words in context, and professional terms).

Summary

With approval from the Seton Hall University Institutional Review Board granted on July 26, 2005 the researcher used two instruments for determining the insights, attitudes and experiences of directors, supervisors, or coordinators of special education programs in public school districts in New Jersey toward the consequences of large-scale assessment results for students with disabilities. A second purpose was to determine if the accountability requirement has had an effect on curriculum, instruction, and special education placement. The first instrument was the PASE survey. The second instrument was a focus group session. It is believed that the results from the sample survey participants and the focus groups participants will result in conclusions that may be generalized to the larger population of directors of special services.

Chapter IV

Research Findings

Overview

The purpose of the study is to explore the directors of special services perceptions of the consequences of high-stakes assessment and its impact on the increased demand for accountability on special education services in 100 public school districts in central and northern New Jersey. This chapter summarizes the results of the data analysis for the research questions posed in this study:

1. What is the perception of the directors of special services toward changes in student achievement and instruction of special education students since the inclusion of students with disabilities in the New Jersey state assessment system?
2. What is the perception of accountability in special education of directors of special services in low socioeconomic districts (DFG Groups) as compared to the perception of accountability in special education of directors of special services in high socioeconomic districts (DFG Groups)?
3. What is the perception of accountability in special education of directors of special services in high-enrollment districts as compared to the perception of accountability in special education of directors of special services in low-enrollment districts?
4. What is the perception of accountability in special education of directors of special services with less experience as compared to the perception of accountability in special education of directors of special services with more experience?

5. What is the of perception of the directors of special services toward changes in use of testing accommodations since the inclusion of students with disabilities in the New Jersey state assessment system?
6. What is the diversity of perceptions of the directors of special services toward changes in professional development since the inclusion of students with disabilities in the New Jersey state assessment system?
7. What is the diversity of perceptions of the directors of special services toward the effectiveness of the Alternate Proficiency Assessment since the implementation of New Jersey's state alternative assessment system for students with severe disabilities?
8. What is the diversity of perceptions of the directors of special services toward the effect of including students with disabilities in the New Jersey statewide assessment program?

Data for this study is presented by methodology. This descriptive study employed two research methodologies to answer the research questions: administering a survey (quantitative design) and conducting a focus group session (qualitative design).

The quantitative findings of the research are presented first. The return rate and comments from the Perceptions of Accountability in Special Education (PASE) surveys are discussed. Overall descriptions of the results are provided.

The qualitative findings were examined according to whether directors of special services viewed accountability as positive, neutral or negative. Table 18 encapsulates this information based on the responses to the questions posed during the focus group sessions.

Quantitative Research Results

According to the New Jersey Department of Education, there are 118 districts in central and northern New Jersey. Of the original 118 districts considered for participation in the study, four participated in the pilot study, 14 districts did not meet the determiner requirements (see previous definition) and were eliminated as possible participants. Therefore, 100 PASE surveys were sent (Attachment A). Of this number, 68 were returned in time for inclusion in this study (return rate of 68%). One survey was not included in the data analysis due to incomplete responses from the participant.

Research Question # 1 Frequencies of Perceptions of Directors of Special Services Toward Changes in Student Achievement and Instruction of Special Education Students Since the Implementation of the New Jersey's State Assessment System

The average response to the PASE survey ranged from an average low score of 1.86 to a high score of 4.43. The breakdown of frequencies is shown in Table 1, clustered as follows: disagree with values 1-2 (21); neutral with a value of 3 (37); and agree with values of 4-5(9). Responses (55.3%) of the fell in the neutral response area.

Table 1

Summary of Frequencies for Changes in Student Achievement and Instruction

| | Frequency | Percentage |
|----------|-----------|------------|
| Disagree | 21 | 31 |
| Neutral | 37 | 55 |
| Agree | 9 | 13 |

This percentage had been rounded and
therefore do not total 100%

The average score for all the PASE survey scores was 3.29, suggesting that there has been no noticeable change in student achievement and instruction for student with disabilities since their inclusion in the New Jersey state assessment system. However,

item analysis indicates that survey questions 4 and 15 pertaining to research question 1 indicates agreement, average score for all the PASE survey scores was 4, indicates that the directors' perception of improvement in the special education teachers' knowledge of the general education curriculum since the inclusion of students with disabilities in the New Jersey Assessment system. Results also indicate that directors agreed, average score for all the PASE survey scores was 4, that other alternative assessment methods than the current system that would accurately assess student progress for students with disabilities. Survey questions 2, 9, and 11 indicate a neutral response; average score for all the PASE survey scores was 3, of how directors perceive an improvement in general education teachers' ability to modify the curriculum when students with disabilities are included in their classes. Directors' answers indicated a neutral response that students with disabilities scores accurately reflect progress toward the NJCCCS when taking the general education assessment or the APA.

Research Question #2 Director's Perception Toward Accountability of Students with Disabilities with Respect to District Factor Group

This question sought to determine if a difference in perceptions existed among directors toward accountability for student with disabilities with respect to district factor group (DFG). For this study, DFG groups were identified as Group AB, Group CD,DE,FG,GH,H, and Group IJ.

The breakdown of responses for the DFG group is provided in Table 2

| | | | | | | | | | |
|---------|--|--|--|--|--|--|--|--|----|
| | | | | | | | | | |
| | | | | | | | | | 77 |
| Table 2 | <i>Summary of Received Surveys with Respect to DFG</i> | | | | | | | | |

| DFG Group | Total Surveys Received |
|-----------|------------------------|
| AB | 9 |
| CDDEFGGHH | 30 |
| IJ | 28 |

The PASE Score and DFG Group's number of cases, mean, standard deviation, standard error, 95 % confidence interval and minimum and maximum, for each dependent variable and independent variables is described in Table 3. The average score for the DFG group responses was 3.29 for the PASE score.

Table 3

Descriptive Statistics for Accountability by DFG Group

PASE SCORE

| | N | M | SD | SE | 95% Confidence Interval for Mean | | Min | Max |
|-----------|----|--------|--------|--------|----------------------------------|-------------|------|------|
| | | | | | Lower Bound | Upper Bound | | |
| AB | 13 | 3.1713 | .34682 | .09619 | 2.9617 | 3.3809 | 2.80 | 3.86 |
| CDDEFGGHH | 29 | 3.3783 | .37182 | .06904 | 3.2368 | 3.5197 | 2.40 | 3.93 |
| IJ | 25 | 3.2659 | .46914 | .09383 | 3.0722 | 3.4595 | 2.47 | 4.20 |
| Total | 67 | 3.2962 | .40858 | .04992 | 3.1965 | 3.3958 | 2.40 | 4.20 |

Hypothesis for question 2. The hypothesis for this question is stated in null form. Ho: The average perception of accountability score is the same for all three district socioeconomic status groups. The formula for the hypothesis ($H_0: \mu_1 = \mu_2 = \mu_3$) states that there is no significant difference in the average perception of accountability between schools from different socioeconomic status groups as defined by the district's DFG Grouping. If the null hypothesis is rejected and H1 is accepted, this would mean that there is a significant difference in a director's perception of accountability between schools from different

| | | | | | |
|--|--|--|--|--|----|
| | | | | | 78 |
| | | | | | |

socio-economic status groups as defined by the district's DFG Grouping. The .05 alpha level of significance was used for determining rejection of Ho and acceptance of H1.

To determine the outcome for this hypothesis, an analysis of variance (ANOVA) was performed on the data and an F value of 1.272 was obtained. Results are reported in Tables 4 and 5.

Table 4

Test of Homogeneity of Variances for Accountability by DFG Group

PASE SCORE

| | | | |
|------------------|-----|-----|------|
| Levene Statistic | df1 | df2 | Sig. |
| 1.437 | 2 | 64 | .245 |

Table 5

ANOVA for Accountability by DFG Group

PASE SCORE

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|----|-------------|-------|------|
| Between Groups | .421 | 2 | .211 | 1.272 | .287 |
| Within Groups | 10.597 | 64 | .166 | | |
| Total | 11.018 | 66 | | | |

Levene's Test of Homogeneity with a significance level of .05 or better was used to test if there were equal variances across samples. Levene's Test of Homogeneity is not significant (.245) indicating the assumption of homogeneity for the variances has not been violated. The one-way ANOVA results for the main effect of DFG Group on accountability as indicated by the average PASE score is not significant with an F value of 1.272 (df=2,64, p=.287). The F value of 1.272 indicates the research fails to reject the null hypothesis and accepts that there is no significant difference in the directors of special services perception of accountability with respect to schools from different DFG

| | |
|--|--|
| | |
| | |
| | |
| | |
| | |

Groups. This implies that DFG Grouping does not have a significant effect on perception of accountability as indicated by the average PASE score.

Research Question #3 Director's Perception Toward Accountability of Students with Disabilities with District Enrollment as per the 2004 ASSA Report

This question sought to determine if a difference in perceptions existed among directors toward accountability for student with disabilities with respect to district enrollment as per the 2004 ASSA Report. For this study the district enrollment groups were defined as: 500-800, 801-1000, 1001-3000, 3001-5000, and more than 500.

The breakdown of responses for the district enrollment groups is provided in Table 6.

Table 6

Summary of Received Surveys with Respect to District Enrollment Groups

| District Enrollment | Total Surveys Received |
|---------------------|------------------------|
| 500-800 | 8 |
| 801-1000 | 7 |
| 1001-3000 | 23 |
| 3001-5000 | 19 |
| 5001 and more | 10 |

The PASE score and District Enrollment Group's number of cases, mean, standard deviation, standard error, 95% confidence interval, and minimum and maximum, for each dependent variable and independent variables is described in Table 7. The average PASE score for the District Enrollment group responses was 3.26.

Table 7

Descriptive Statistics for Accountability by District Enrollment Group

PASE SCORE

| | N | M | SD | SE | 95% Confidence Interval for Mean | | Min | Max |
|---------------|----|--------|--------|--------|-------------------------------------|----------------|------|------|
| | | | | | Lower Bound | Upper Bound | | |
| 500-800 | 8 | 3.2500 | .52705 | .18634 | 2.8094 | 3.6906 | 2.53 | 3.93 |
| 801-1000 | 7 | 3.5048 | .34610 | .13082 | 3.1847 | 3.8249 | 2.87 | 3.93 |
| 1001- 3000 | 23 | 3.2812 | .42733 | .08910 | 3.0964 | 3.4659 | 2.40 | 4.20 |
| 3001- 5000 | 19 | 3.3228 | .42703 | .09797 | 3.1170 | 3.5286 | 2.47 | 4.00 |
| 5001- 9000 | 10 | 2.9600 | .29178 | .09227 | 2.7513 | 3.1687 | 2.60 | 3.47 |
| Total | 67 | 3.2647 | .42876 | .05238 | 3.1601 | 3.3693 | 2.40 | 4.20 |

Hypothesis for question 3. The hypothesis for this question is stated in null form. H_0 :

The average perception of accountability score is the same for all five district size groups.

The formula for the hypothesis ($H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$) states that there is no significant difference in the average perception of accountability between schools of different enrollment sizes. If the null hypothesis is rejected and H_1 is accepted, this would mean

that there is a significant difference in a director's perception of accountability between schools from different enrollment groups as defined by 2004 ASSA report. The .05 alpha level of significance was used for determining rejection of H_0 and acceptance of H_1 .

To determine the outcome for this hypothesis, an analysis of variance (ANOVA) was performed on the data and an F value of 2.028 was obtained. Results are reported in Tables 8 and 9.

Table 8

Test of Homogeneity of Variances for District Enrollment Group

| PASE SCORE | | | | | |
|------------|------------------|-----|-----|------|--|
| | Levene Statistic | df1 | df2 | Sig. | |
| | 1.026 | | | .401 | |

Table 9

ANOVA for District Enrollment Group

| PASE SCORE | | | | | |
|----------------|----------------|----|-------------|-------|------|
| | Sum of Squares | df | Mean Square | F | Sig. |
| Between Groups | 1.404 | 4 | .351 | 2.028 | .101 |
| Within Groups | 10.729 | 62 | .173 | | |
| Total | 12.133 | 66 | | | |

Levene's Test of Homogeneity with a significance level of .05 or better was used to test if there were equal variances across samples. Levene's Test of Homogeneity is not significant (.401) indicating the assumption of homogeneity for the variances has not been violated. The one-way ANOVA results for the main effect of district size on accountability as indicated by the average PASE score is not significant with an F value of 2.028 (df=4, 62, p=.101). The F value of 2.028 indicates the research fails to reject the null hypothesis and accepts that there is no significant difference in a director of special services perception of accountability with respect to schools of different enrollment sizes. This implies that district size does not have a significant effect on the perception of accountability as indicated by the average PASE score.

| | |
|--|--|
| | |
| | |
| | |

Research Question #4 Director's Perception Toward Accountability of Students with Disabilities with Respect Years of Experience in the Position of Director of Special Services

This question sought to determine if a difference in perceptions existed among directors toward accountability for student with disabilities with respect to years of experience in the position of director of special services. For this study the groups were defined as less than 1 year, 1-3 years, 4-9 years, and 10 or more years.

The breakdown of responses for the years for the years of experience in the position of director of special services group is provided in Table 10.

Table 10

Summary of Received Surveys with Respect to Years of Experience

| Years Of Experience | Total Surveys Received |
|---------------------|------------------------|
| 1-3 years | 12 |
| 4-9 years | 27 |
| 10 or more years | 28 |

The PASE score and Years of Experience group's number of cases, mean, standard deviation, standard error, 95% confidence interval, and minimum and maximum, for each dependent variable and independent variables is described in Table 11. The average PASE score for the Years of Experience group responses was 3.26.

Table 11

Descriptive Statistics for Accountability by Years of Experiences

PASE SCORE

| | N | M | SD | SE | 95% Confidence Interval for Mean | | Min | Max |
|-------------------|----|--------|--------|--------|-------------------------------------|----------------|------|------|
| | | | | | Lower Bound | Upper Bound | | |
| 1-3 YEARS | 12 | 3.2500 | .41500 | .11980 | 2.9863 | 3.5137 | 2.47 | 3.80 |
| 4-9YEARS | 27 | 3.2790 | .45179 | .08695 | 3.1003 | 3.4577 | 2.40 | 4.20 |
| 10ORMORE YEARS | 28 | 3.3595 | .33913 | .06409 | 3.2280 | 3.4910 | 2.80 | 3.93 |
| Total | 67 | 3.3075 | .39780 | .04860 | 3.2104 | 3.4045 | 2.40 | 4.20 |

Hypothesis for question 4. The hypothesis for this question is stated in null form. H_0 : The average perception of accountability score is the same for all four directors of special services experience level groups. The formula for the hypothesis ($H_0: \mu_1 = \mu_2 = \mu_3$) states that there is no significant difference in the average perception of accountability between schools of different enrollment sizes. If the null hypothesis is rejected and H_1 is accepted, this would mean that there is a significant difference in a director's perception of accountability between director's years of experience in the position of director of special services. The .05 alpha level of significance was used for determining rejection of H_0 and acceptance of H_1 .

To determine the outcome for this hypothesis, an analysis of variance (ANOVA) was performed on the data and an F value of .426 was obtained. Results are reported in Tables 12 and 13.

Table 12

Test of Homogeneity of Variances for Years of Experience

| PASE SCORE | Levene Statistic | df1 | df2 | Sig. |
|------------|------------------|-----|-----|------|
| | 1.359 | 2 | 64 | .264 |

Table 13

ANOVA of Years of Experience

| PASE SCORE | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|----|-------------|------|------|
| Between Groups | .137 | 2 | .069 | .426 | .655 |
| Within Groups | 10.307 | 64 | .161 | | |
| Total | 10.444 | 66 | | | |

Levene's Test of Homogeneity of variances with a significance level of .05 or better was used to test if there were equal variances across samples. Levene's Test is not significant (.264) indicating the assumption of homogeneity for the variances has not been violated. The one-way ANOVA results for the main effect of years of experience in the position of director on accountability as indicated by the average PASE score is not significant with an F value of .426 (df=2, 64, p=.655). The F value of .426 indicates the research fails to reject the null hypothesis and accepts that there is no significant difference in a director of special services perception of accountability with respect to years of experience in the position of director. This implies that years of experience in the position of director does not have a significant effect on the perception of accountability as indicated by the average PASE score.

Research Question# 5-Frequencies of Perceptions of the Directors of Special Services Toward Changes in Use of Testing Accommodations Since the Implementation of New Jersey's State Assessment System

The average response to the PASE survey ranged from an average low score of 2 to a high score of 5. The breakdown of frequencies is shown in Table 14, clustered as follows: disagree with value 2 (10), neutral with a value of 3 (11), and agree with values of 4-5(46). Responses (68.6%) fell in the agree response area.

Table 14

Summary of Frequencies with Respect to Changes in Use of Testing Accommodations

| | Frequency | Percentage |
|----------|-----------|------------|
| Disagree | 10 | 15% |
| Neutral | 11 | 16% |
| Agree | 46 | 69% |

The average score for all the PASE survey scores was 3.67 suggesting that there has been no noticeable change in curricular accommodations with the adaptations continuing to be used for instruction and testing.

Research Question # 6- Frequencies of Perceptions of Directors of Special Services Toward Changes in Professional Development Since the Implementation of the New Jersey's State Assessment System

The average response to the PASE survey ranged from an average low score of 2 to a high score of 5. The breakdown of frequencies is shown in Table 15, clustered as follows: disagree with value 2 (21), neutral with a value of 3 (14), and agree with values of 4-5(32). Responses (47.8%) fell in the agree response area.

Table 15

Summary of Frequencies with Respect to Professional Development

| | Frequency | Percentage |
|----------|-----------|------------|
| Disagree | 21 | 31 |
| Neutral | 14 | 21 |
| Agree | 32 | 48 |

The average score for all the PASE survey scores was 3.22 suggesting that there has been some noticeable change in professional development since the inclusion of student with disabilities in the New Jersey state assessment system but the effect of the professional development is unknown at this time.

Research Question# 7 Frequencies of Perceptions of the Directors of Special Services Toward the Effectiveness of the Alternate Proficiency Assessment since the Implementation of New Jersey's State Assessment System for Students with Severe Disabilities

The average response to the PASE survey ranged from an average low score of 1.25 to a high score of 3.25. The breakdown of frequencies is shown in Table 16, clustered as follows: disagree with value 2 (66), neutral with a value of 3 (1), and agree with values of 4-5(0). Responses (98.5%) fell in the response area of disagree.

Table 16

Summary of Frequencies with Respect to the Effectiveness of the Alternate Proficiency Assessment

| | Frequency | Percentage |
|----------|-----------|------------|
| Disagree | 66 | 98 |
| Neutral | 1 | 2 |
| Agree | 0 | 0 |

The average score for all the PASE survey scores was 2.21 suggesting that effectiveness of the Alternate Proficiency Assessment (APA) is minimal since the

implementation of New Jersey's state assessment system for students with severe disabilities. Itemized analysis of the survey questions 8, 11, 12 indicates a neutral response with an average score of 3, that APA scores may distort the districts adequate yearly progress (AYP); that the APA scores accurately reflect student progress toward the New Jersey Core Curriculum Content Standards (NJCCCS); and that the APA measures program quality more than student progress.

| | | |
|--|--|--|
| | | |
| | | |
| | | |

Research Question # 8 Frequencies of Perceptions of the Directors of Special Services Toward the Effect Inclusion Since the Implementation of the New Jersey's State Assessment System

The average response to the PASE survey ranged from an average low score of 1 to a high score of 5. The breakdown of frequencies with respect to the effectiveness of inclusion in testing is shown in Table 17, clustered as follows: disagree with value 2 (22), neutral with a value of 3 (10), and agree with values of 4-5(35). Responses (52%) fell in the agree response area.

Table 17

Summary of Frequencies with Respect to the Effectiveness of Inclusion in Testing

| | Frequency | Percentage |
|----------|-----------|------------|
| Disagree | 22 | 33 |
| Neutral | 10 | 15 |
| Agree | 35 | 52 |

The average score for all the PASE survey scores was 3.20, suggesting that the New Jersey's state assessment system has had a somewhat positive effect on the inclusion of students with disabilities in general education classes.

Summary of Quantitative Research Results

Quantitative data results indicate that directors of special services:

1. Perceive assessment to have a positive effect on special education services;
2. Indicate that student achievement and learning has improved since the inclusion of students with disabilities in the New Jersey Assessment process;
3. Indicate that professional development opportunities for special education teachers and general education teachers has improved since the inclusion of students with disabilities in the New Jersey Assessment process; and
4. Indicate that there was no significant difference between responses based on the district factor group, district enrollment, or the years of in the position of director of special services.

Qualitative Research Results

A focus group was held as part of the qualitative design for this study. The focus group was employed in an effort to provide quality information to New Jersey's school districts regarding the consequences of standardized assessment on students with disabilities. It is believed that the introduction of this technique would enhance the understanding of the quantitative survey data.

The questions for the focus group were developed by the researcher and determined and sequenced based on an analysis of the literature, focus group questions from previous doctoral dissertations, and based on open-ended survey questions from previous National Council for Educational Outcome research (Attachment B). The question guide was arranged in a logical manner to facilitate group discussion, and each type of question was designed for a specific purpose. A second jury of experts was asked

to review the proposed questions for the focus group. Four directors of special education from preschool through Grade 12 districts were asked to determine that the questions demonstrated face and content validity. Questions were open-ended and non-specific so that the participant was able to interpret the questions based on his or her own experiences and follow the principles of qualitative interviewing to assure construct validity. Participants had the opportunity to verify the summary statement posed at the end of the session.

The composition of the focus group was similar to the survey participants. Participants of the focus group were directors of special services who have direct contact with the members of IEP team, including parents, case managers, general education and special education teachers. The directors are also responsible for monitoring curriculum and instructional implementation and reporting and analyzing reports and scores. The focus group was comprised of participants from K-8 and K-12 districts. The group was homogeneous in nature with respect to certification, but heterogeneous with respect to years of experience in education, district enrollment and DFG group.

There were 12 questions:

1. What evidence of change have you observed and experienced since the implementation of the New Jersey Assessment system relative to students with disabilities?
2. What do you see as a positive result of the participation of students with disabilities in the assessment and accountability system?
3. What do you see as a negative consequence of the participation of students with disabilities in the assessment and accountability system?

4. An intended primary outcome of standards-based reform is improved student learning through the establishment of high standards for all students and assessment of student progress toward meeting these standards. How does the NJ Assessment system do this for students with disabilities?

5. What impact has the NJ Assessment program and the inclusion of students with disabilities had on child study teams and teachers in your districts?

6. What impact, if any, has the NJ Assessment program and the inclusion of students with disabilities had on the students?

7. What have the parents of your students said about the NJ assessment system?

8. What impact has the inclusion of students with disabilities in the NJ Assessment process had on your district's accountability status?

9. What approach does your district use to determine which students participate in the alternate assessment?

10. How are accommodations determined in your district?

11. What are some of the major obstacles to including students with disabilities in the district accountability system?

12. Describe any other methods of assessment you believe would meet the assessment requirements?

The research questions became the basis for analyzing the qualitative data, with the focus of the questioning identifying accountability.

Comments of the directors of special services were coded according to themes and patterns. The researcher's professional judgment was used in the coding based on an analysis of words used in the focus groups (word repetitions, key-words in context, and

professional terms). The comments were coded based on the directors of special services perception of accountability, student learning, and professional development as having a positive, neutral, or negative effect. Table 18 encapsulates this information based on the 12 questions posed during the focus group session.

Table 18

Summary of Focus Group Responses

| Parent Reaction | Positive effect | Neutral effect | Negative effect | No comment |
|--|-----------------|----------------|-----------------|------------|
| Student Learning and Instructional Delivery | 10 | 2 | | 4 |
| Professional Development | 8 | 4 | 3 | 1 |
| Effectiveness of APA for assessment purposes | | 4 | 10 | 2 |
| Participation in General Testing | 4 | 6 | 1 | 5 |
| Effect of assessment on District AYP | | 10 | 3 | 3 |
| Accommodations | 5 | 7 | | 4 |
| Parent Reaction | | 5 | 6 | 5 |
| Impact on Inclusion | 3 | 5 | | 8 |

The comments made during the focus group session highlighted the overall perception of accountability of special education students. The directors often referenced specific concerns that existed within their district. Typically these concerns were related to federal or state government regulations and differences in philosophical opinion.

General comments and observations reflect the greatest impact of accountability on special education services is in the areas of instruction, student learning, and professional development. Directors of Special Services report:

1. there is an enhanced relationship between instruction in special education classes and the New Jersey Core Curriculum Content Standards (NJCCCS);

2. there is a decrease in altering the instructional materials; instead the learning expectations are being modified, but the same materials are used during instruction for regular education and special education;

3. replacement classes have been added to the curriculum to provide tutoring for students scoring not proficient or in danger of failing the state tests;

4. increased awareness of the importance of using accommodations throughout the day to day instruction and not just during testing situations;

5. instructional time in self-contained special education classes is compromised during the final weeks of the Alternate Proficiency Assessment (APA) to allow for preparation of the portfolio;

6. there has been an increase in professional development opportunities for general education teachers to include training on inclusion and instructional strategies and techniques, and;

7. allowance of professional development days for gathering of information and

compiling the portfolio materials for the APA

A second major theme for comments and observations focused on the APA and its impact on special education services. Directors of Special Services report:

1. it is difficult to establish consistency of testing procedures from year to year due to teacher turn-over;
2. it is difficulty to conduct item analysis to assess successes and failures due to the subjectivity of the assessment;
3. APA is not a comprehensive assessment of student progress since it only assesses three curricular areas and four goals from each area; and
4. although the purpose of the assessment is to improve student learning, the amount of time and attention invested in the preparation and presentation of the portfolio contradicts the purpose of the assessment.

Summary

Data gathered in this study used both quantitative and qualitative analysis to determine the directors of special services perceptions of the consequences of high-stakes assessment and its impact on the increased demand for accountability on special education services. The data presented in this chapter illustrated that directors of special services do perceive assessment to have a positive effect on special education services. Specifically student achievement and learning and professional development within the district has improved since the inclusion of students with disabilities in the New Jersey Assessment process.

There was no significant difference between responses based on the district factor group, district enrollment, or the years of in the position of director of special services. The qualitative data generated by the focus group provided a broad, comprehensive view of the directors' perceptions of accountability and its impact on special education services. Student achievement and instructional delivery and professional development are indicated to have been positively affected while effectiveness of the APA to assess student progress relative to the NJCCCS are indicated to have had a negative impact. These findings expand the research in the area of assessment and students with disabilities and provided a compilation of research that can be further explored within the context of accountability.

Chapter V

Summary

Knowing the extent to which all students are learning, is used to define educational policy, district funding, curricular decisions, and student class placement. As a means to assess progress, states have implemented large-scale, high-stakes testing. Local school districts are held accountable for student progress by the reporting of test scores.

Federal policy and the quest for improved schools have elevated the importance of assessment results for the purpose of school improvement and student achievement for all students, including students with disabilities. The IASA, IDEA '97, IDEA 2004 and NCLB 2001 are the most recent legislation to impact the educational reform movement that considers the learning needs of all students. Each of these laws has the common requirement that students with disabilities be included in the assessment system.

As a result of this legislation, the New Jersey Department of Education and local school districts had to confront what it means for all students to achieve at high standards. This indicates a policy shift away from the special education ideal of students meeting individualized goals to the inclusion of their progress into a standardized system of accountability. Directors of special services are faced with challenges of merging the ideals of special education in to the one size fits all general education system. They must lead their district in the policy shift to assure special education students are appropriately and accurately assessed.

— Studies have been cited in this dissertation of classroom teachers' perceptions of the alternate assessment process and of state directors of special education related to issues of accountability and assessment reliability and validity. This research study was designed to explore the consequences of large-scale high-stakes assessment for students with disabilities through the point of view, experiences and perceptions of the directors of special services in school districts in central and northern New Jersey. Directors of special services, being the primary administrator responsible for special education student placement and progress, have a valid insight into the usefulness of New Jersey's assessment system and its impact on special education services. The study attempted to collect and describe the directors' perceptions of the impact of the New Jersey Assessment system on special education services.

The methodology for the study included both quantitative and qualitative measures. To answer the research questions, a survey entitled "Perceptions of Accountability Special Education" (PASE) survey was created. A first draft survey was sent to a jury of experts who critiqued the contents and made suggestions. The survey was amended and sent to a pilot group for further refinement. Out of the 100 surveys sent, 67 useable surveys were returned.

Descriptive statistical data regarding overall directors of special services perceptions of accountability and its impact on special education services was provided by this study. The dependent variable was the average score on the PASE survey. Directors' perceptions were measured for the dependent variable based on like responses on the Likert scale instrument.

All three independent variables were categorical. They included District Factor Group which is an indicator of socioeconomic status, district enrollment, and years of experience in the position of director of special services.

Three null hypothesis were developed to guide the statistical analysis of the study.

These included:

1. There is no significant difference in the average perception of accountability score is the same for all three district socioeconomic status groups.
2. There is no significant difference in the average perception of accountability score is the same for all five district size groups.
3. There is no significant difference in the average perception of accountability score is the same for all four directors of special services experience level groups.

Research questions quantified the directors' perceptions concerning changes in student achievement and instruction, use of testing accommodations, changes in professional development, the effectiveness of the Alternate Proficiency Assessment (APA), and the effect of including students with disabilities in the New Jersey assessment process.

Certain assumptions were inherent to this study. It was assumed that directors responded honestly to anonymous survey. Further it was assumed the focus group participants were honest in sharing their opinions of accountability and its impact on special education services.

As with any research study, limitations exist that may pose a threat to finding answers to the research questions. Survey research is typically strong on reliability but weak on validity. To enhance the validity, the survey was piloted with directors of special services. The research methods were appropriate for the intent of this study, but it is

important to note that caution must be taken when interpreting and attempting to generalize the findings on a state level or between states. This study was not designed to show causal relationships between the consequences of large-scale high-stakes assessment and specific educational outcomes for students with disabilities.

A qualitative research methodology was employed to help control limitations and the inability of the directors to provide open-ended responses. A focus group was used in an attempt to place the survey results within the proper context. However, it must be noted that focus group participants represented a sample population of the directors of special services in central and northern New Jersey.

Findings

The data analysis revealed that directors of special services do perceive assessment and the need to be held accountable for student progress to have a neutral to positive effect on special education services. Specifically student achievement and professional development within the district has improved since the inclusion of students with disabilities in the New Jersey Assessment process.

With respect to student achievement and learning, directors perceived an improvement in the special education teachers' knowledge of the general education curriculum since the inclusion of students with disabilities in the New Jersey Assessment system. Focus group participants commented that there is an enhanced relationship between instruction in special education classes and the New Jersey Core Curriculum Content Standards (NJCCCS) and a decrease in the altering of the instructional materials. Instead the learning expectations are being modified, but the same materials are used

during instruction for regular education and special education. However caution should be used when determining if the assessment process has had a direct effect on student achievement since the study was not designed to show a causal relationship between the consequences of high-stakes testing and student achievement.

Another perceived benefit of the accountability mandate is the development of replacement classes, which have been added to the curriculum to provide tutoring for students scoring not proficient or in danger of failing the state tests.

Survey results indicate there was no change in the curriculum modifications by the general education teachers' when students with disabilities are included in their classes. Since the modifications are guided by the IEP they were consistently implemented. The survey and focus group participants were in agreement that curricular accommodations are consistently used for instruction and testing since the implementation of the New Jersey's state assessment system. An increased awareness of the importance of using accommodations throughout the day-to-day instruction and not just during testing situations has become evident with teachers since the inclusion of students with disabilities in the assessment process.

Directors also perceived that students with disabilities scores accurately reflect progress toward the NJCCCS when taking the general education assessment and the Alternate Proficiency Assessment (APA).

Directors also perceive that another assessment system other than the one that is currently being used would more accurately assess student progress for students with disabilities. General comments and observations reflect the greatest impact of

accountability on special education services is in the areas of instruction, student learning, and professional development.

There was agreement in the finding that effectiveness of the Alternate Proficiency Assessment (APA) is minimal since the implementation of New Jersey's state assessment system for students with severe disabilities. Survey results and focus group responses indicate that APA scores distort the district's Adequate Yearly Progress (AYP) and that the APA measures program quality more than student progress. Survey participants and focus group participants indicated a positive effect of the APA is that the APA scores accurately reflect student progress toward the NJCCCS. However, the focus group participants stated that the APA is limiting in noting student progress. This is due to the assessment being based on the teacher selection to assess the student on two language arts goals, two math goals, and three science goals. The directors stated the assessment does not accurately reflect the overall student learning as compared to the comprehensive general assessment.

Students are being included in regular education classes more frequently than before the assessment mandate. Caution should be used, however, when concluding that the assessment mandate has resulted in more students with disabilities being included in regular classrooms. A factor influencing this movement may be the mandate for inclusion and a move away from self-contained classes.

Survey respondents and focus group participants perceived an improvement in professional development. Data from the focus group indicates that there has been an increase in professional development opportunities for general education teachers to

include training on inclusion and instructional strategies and techniques. There is an increased awareness of the NJCCCS for special education teachers.

The hypotheses developed for the research were supported by the data. Specifically the hypothesis that the average perception of accountability score is the same for all 3 district socioeconomic status groups was accepted implying that the socioeconomic group as represented by the DFG Grouping does not have a significant effect on perception of accountability as indicated by the average PASE score.

The remaining hypotheses were not supported by the data. The district size or years of experience of the director of special services does not have a significant effect on the perception of accountability as indicated by the average PASE score.

Implications

The findings of this study, dealing with the directors of special services perceived impact of accountability on special education services has corroborated the findings in the literature: The impact of accountability in special education has improved the instructional process for students with disabilities. School districts are now held accountable to assure students with disabilities are held to the same academic standards as all other students resulting in higher expectations for the students with disabilities. There is an increased participation in assessment and accountability measures and there is improved instruction. Professional development opportunities have expanded for general education teachers to include training on inclusion and instructional strategies and techniques for students with disabilities. It is important to remember, however, that the inclusion of students with disabilities in high-stakes large-scale assessment is in its

infancy and there are numerous pathways of inquiry that need to be considered. The recommendations made here will be limited to major summary points.

The inclusion of students with disabilities is viewed with mixed results. While it appears unlikely that educators and parents would argue that in the past that there have been low achievement expectations for students with disabilities, we still do not know the full impact of including students in high-stakes large-scale testing. The results of this study indicate some positive findings and some areas that continue to be of great concern. As found in the literature, systematic concerns related to the reliability and validity of the alternate assessment process exist. Questions remain on whether assessment meets the desired outcome of measuring student performance. Results indicated that the directors believe the assessment measures program quality more than student achievement. Directors also believed that instructional time is compromised due to the large amount of time allocated for the preparation of the portfolio.

Recommendations

Future Research

Future research should involve the collection of longitudinal data to ascertain if there is an association between the inclusion of students in high-stakes large-scale testing and student achievement as evidenced by meeting state standards. Research should investigate the effects of students with disabilities being required to achieve at the same grade level standards as non-disabled students. Accountability systems should be

examined to look at the impact of student scores on planning and preparation, the classroom environment, instruction, and professional development.

Further research should be conducted on the use of the alternate assessment. Rigorous standards must be established that assure the validity and reliability of the alternate assessment process. The assessment must be developed or refined to accurately reflect student progress in all areas not just a few selected goals. The assessment process must become more objective and comprehensive in its evaluation of student performance. While it is important to assess program quality and performance more weight should be placed on student achievement than the program components. The process must also become user friendly and time efficient. Until these criteria can be established, questions will continue to remain about the appropriateness and benefits of the alternate assessment process.

Although the directors of special services have direct knowledge of policy and curriculum, their perceptions may not accurately reflect the implementation concerns related to the assessment process. Therefore, conducting future research with classroom personnel, specifically the classroom teachers, may be helpful when comparing results related to improvement in instruction and student learning.

Future surveys should include additional demographic criteria. Understanding the types of disabilities served within the director of special services district would provide additional data on the effects of the assessment process relative to the type of disability.

State and National Policy Implications

Students' inclusion in the accountability process helps to assure their successful attainment of higher educational standards. The practice of using assessment scores for students with disabilities when determining school resources and rewards and sanctions needs to be examined in light of the need for equity, consistency, and effectiveness for schools with a diverse student population. The question remains, should students with disabilities testing results have the same impact on the final AYP as the performance of other students.

Included in the efforts to improve outcomes for students with disabilities is professional development. States and local districts must support the teachers in developing a high quality assessment of skills that is authentic if the alternate assessment if going to improve the outcomes of students with disabilities. Professional development opportunities should be examined to assure teachers have a clear understanding of the content for which the students are being held accountable. Sustainability factors for a quality professional development plan include a deliberate plan for training, coaching and modeling, opportunities for teachers to understand how the assessment fits their students, the availability of peer networks and support and collegial learning opportunities, and administrative support.

Pre-service programs should be examined to ensure strong curricular content knowledge by the graduates with degrees in special education. Reorganization of undergraduate coursework may have to be considered to assure the graduates are prepared for the demands of instruction in special education.

Content standards that are restricted to academics and do not include a wide range of curricular options for students with disabilities should be examined. If the primary goal of the standards is to provide instruction for students that will prepare them for full participation in the community as an adult, then the standards need to reflect this goal for all students, not just students along an academic path. There is a critical need to design standards and assess students in a rigorous manner but in a manner that is applicable for all learners.

Reporting of assessment performance data is critical to true educational reform. States and local districts must establish a process for sharing the data and then use the information to improve outcomes and achievement level for all students. States should provide data from all test takers whether they participate with or without accommodations or use an alternate assessment. Records should be maintained on the use of accommodations during the test and on a day-to-day basis to assure consistency of use. The reporting process must be clear, concise, and readily understood if results are to be interpreted and useful. The data should be consistent between states to provide comparative information about schools, districts, and states.

Summary

The results of the current study indicate some positive findings and some areas that continue to be of great concern. There is an increased participation in assessment and accountability measures and there is improved instruction. Professional development opportunities have expanded for general education teachers to include training on

inclusion and instructional strategies and techniques for students with disabilities. As found in the literature, a systematic concern related to the reliability and validity of the alternate assessment process exists. Questions remain to the effectiveness of the Alternate Proficiency Assessment and its ability to assess student progress toward the New Jersey Core Curriculum Content Standards. Results indicated that the directors believe the assessment measures program quality more than student achievement. Directors also believed that instructional time is compromised due to the large amount of time allocated for the preparation of the portfolio. The use of assessment is a key element in educational reform. Longitudinal accountability research needs to be conducted to ascertain if the current assessment process is sustainable. Further research should be conducted to determine if the current perceived improvements in instructional and professional development will have a significant and lasting effect on student learning and instruction.

References

- Albrecht, S., & Joles, C. (Winter, 2003). Accountability and access to opportunity: Mutually exclusive tenets under a hi-stakes testing mandate. *Preventing School Failure*, 47(2), 86-91.
- Allbritten, D., Mainzer, R., & Ziegler, D. (Jan/Feb 2004). Will students with disabilities be scapegoats for school failures? *Teaching Exceptional Children*, 36(3), 74-75.
- Almond, P., Quenemoen, R., Olsen, K., Thurlow, M. (2000) Gray areas of assessment systems (Synthesis Report 32). Retrieved March 4, 2004, from University of Minnesota, National Center of Educational Outcomes.
<http://education.umn.edu/NCEO/OnlinePubs/Synthesis32>
- Amerin, A. L., & Berliner, D.C. (2002). High-stakes testing, uncertainty, and students learning. *Educational Policy Analysis Archives*, 10(18), Retrieved March 16, 2004, from <http://epaa.asu.edu/eppa/v10n18>
- American Educational Research Association (2000,July). AERA positioning statement concerning high-stakes testing in Pre-K12 education. Retrieved March 5, 2004, from <http://www.aera.net/about/policy/stakes.htm>
- Anderson, G. (1999). *Fundamentals of educational research*. New York: The Palmer Press.

- Babbie, E. (1999) *The basics of social research*. Belmont, CO: Wadsworth.
- Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.
- Browder, D. M. (2001). *Curriculum and assessment for students with moderate and severe disabilities*. New York: The Guilford Press.
- Browder, D.M., Fallin, K., Davis, S., & Karvonen, M. (2003). Consideration of what may influence student outcomes on alternate assessment. *Education and Training in Developmental Disabilities, 38*, 255-270.
- Browder, D.M., Spooner, F., Algozzine, R., Ahlgrim-Delzell, L., & Karvonen, M. (2003). What we know and need to know about alternate assessment. *Exceptional Children, 70* (1), 45-61.
- Chudowsky, N., & Behuniak, P. (1998, Winter). Using focus groups to examine the consequential aspect of validity. *Educational Measurement: Issues and Practice, 17*(4), 28-38.
- Cohen, M., & Heumann, J. (2001, January 12). Clarification on the role of the IEP even selecting individual accommodations, modifications in administration, and alternate assessments for state and district-wide assessment of student achievement. Memorandum from the United States Department of Education.
- Coleman, A., Palmer S. (January 2004). Review of final NCLB 2001 rules for severely disabled students. *Special Education Report, 6-8*.

Educational Policy Research Reform Institute (EPRRI), (Spring 2002). Policy updates, issue one. College Park: University of Maryland, Educational Policy Research Reform Institute, The Institute for the Study of Exceptional Children and Youth. Retrieved 11-1-04, from www.epri.org

Elliott, J.L., Erickson, R.N., Thurlow, M., Shriner, J. (2000). State-level accountability for performance of students with disabilities: Five years of change. *The Journal of Special Education*, 34(1), p. 39-47.

Fink, A. (1995). *The survey handbook*. Thousand Oaks, CA: Sage.

Fowler, F.J. (1993). *Survey research methods* (2nd ed.). Newbury Park, CA: Sage.

Fraenkel, J.R. (1996) *How to design research in education*. New York: McGraw-Hill.

Gersten, R., Chard, D., & Baker, S. (2000). Factors enhancing sustained use of research-based instruction practices. *Journal of Learning Disabilities*. 33(5), 445-447.

Improving America's Schools Act (IASA) 1994. Retrieved March 25, 2004 from <http://www.ed.gov/legislation/ESEA/index.html>

Individuals with Disabilities Act (IDEA), P.L. 102-119, 20 U.S.C. 1400et.seq. (1997). Retrieved March 25, 2004 from http://www.ed.gov/offices/OSERS/Policy/IDEA/the_law.html

- Johnson, E., & Arnold, N. (2004). Validating an alternate assessment. *Remedial and Special Education, 25*(5), 266-275. Retrieved January 18, 2005, from <http://web3.epnet.com/externalframe>
- Kampfner, S. H., Horvath, L. S., Kleinert, H. L., & Kearns, J. F. (2001). Teachers' perceptions of one state's alternate assessment: Implications for practice and preparation. *Exceptional Children, 67*, 361-374.
- Kleinert, H.L., & Kearns, J. F. (2001). *Alternate assessment: Measuring outcomes and supports for students with disabilities*. Baltimore: Paul H. Brookes Publishing.
- Kleinert, H.L., Kennedy, S., & Kearns, J. F. (1999). The impact of alternate assessment: a statewide teacher's survey. *The Journal of Special Education, 33*, 93-102.
- Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont Portfolio Assessment Program: Findings and implications. *Educational Measurement: Issues and Practice, 13* (3), 5-16.
- Krueger, R. (2000). *Focus groups: A practical guide for applied research*. Thousand Oaks, CA: Sage.
- Lane, S., Parke, C.S. & Stone, C.A. (1998, Summer). A framework for evaluating the consequences of assessment programs. *Educational Measurement: Issues and Practice, 24-28*.

- Langenfeld, K., Thurlow, M., Scott, D. (1997, January). High-stakes testing for students: unanswered questions and implications for students with disabilities. National Center on Educational Outcomes. Retrieved March 18, 2004, from University of Minnesota, National Center of Educational Outcomes.
http://education.umn.edu/NCEO/OnlinePubs/NCLB_2001_disabilities.htm
- LaMorte, M. W. (2002) *School law*. (7th ed.). Boston: Allyn and Bacon.
- Lawmakers Laud Final Regulations for Testing Disabled. (January 2004). *Education USA*, 8.
- Leddy, P.D. (1997) *Practical research planning and design*. (6th ed.). Upper Saddle River, NJ: Prentice Hall.
- Lehr, C., & Thurlow, M. (2003, October). Putting it all together: Including students with disabilities in assessment and accountability systems. National Center on Educational Outcomes. Retrieved April 22, 2004, from University of Minnesota, National Center of Educational Outcomes
<http://education.umn.edu/nceo/OnlinePubs/Policy16.htm>
- Mann, F. (1970). Human relation skills in social research. In W.J. Filstead, (Ed.). *Qualitative Sociology: Firsthand Involvement with the Social World*. Chicago: Markham Publishing.
- Meherns, W. (1998). Consequences of assessment: What is the evidence? *Educational Policy Analysis Archives* 6, 13.

Messick, S. (1989). Validity. In R.L.Linn (Ed.) *Educational Measurement* (3rd ed.), New York: American Council on Education.

Miller, M.D., & Linn, R.L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement*, 2, 367-378.

Murphy, J. (1980). *Getting the facts: A fieldwork guide for evaluators and policy analysis*. Santa Monica, CA: Goodyear.

New Jersey Alternate Proficiency Assessment 2004: Executive Summary. Retrieved February 1, 2006, from http://www.nj.gov/njded/assessment/apa/2004apa_executives_summary.pdf

No Child Left Behind Act of 2001, 107 U.S.C., P.L. 107-110 (2002). Retrieved February 18, 2004 from <http://www.ed.gov/policy/elsec/leg/esea02/index.html>

No Child Left Behind has Major Implications for Special Education. (2002, November/December). *CEC Today*, 9, 4.

Patton, M.Q. (1990). *Qualitative evaluation and research methods* (2nd ed.) Newbury Park, CA: Sage.

President's Commission on Excellence in Special Education. (2002). *A new era: Revitalizing special education for children and families*. Washington, DC: U.S. Department of Education Office of Special Education and Rehabilitative Services.

- Quenemon, 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, MN: University of Minnesota, National Center of Educational Outcomes. Retrieved 1-11-04 from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis43>
- Rea, L.M., & Parker, R.A. (1997). *Designing and conducting survey research: A comprehensive guide*. San Francisco, CA: Jossey-Bass.
- Roach, A. T., Elliott, S. N., Webb, N. L. (2005). Alignment of an alternate assessment with state academic standards: Evidence for the content validity of the Wisconsin alternate assessment. *The Journal of Special Education*, 38(4), 218-231.
- Rossi, P.H., Wright, J.D., & Anderson, A.B. (1983). *Handbook of survey research*. New York: Academic Press.
- Salvia, J., & Ysseldyke, J.E. (2001). *Assessment (8th edition)*. Boston: Houghton Mifflin.
- Special Needs, Common Goals. (January 8, 2002). *Education Week*, 28(17), 7.
- Thompson, S., Quenemoen, R., Thurlow, M., Ysseldyke, J. (2003). *Alternate assessments for students with disabilities*. Thousand Oaks: Corwin Press, Inc.
- Thompson, S., & Thurlow, M. (2001). 2001 State special education outcomes: A report on state activities at the beginning of a new decade. Retrieved March 18, 2004 from University of Minnesota, National Center of Educational Outcomes

<http://education.umn.edu/NCEO/OnlinePubs/htm>

Thurlow, M.L., Bielinski, J, Minnema, J., & Scott, J. (2002). Out-of-level testing revisited: New Concerns in the era of standards-based reform. In G. Tindal & T. M. Haladyna, (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* pp. 1-26. Mahwah, NJ: Erlbaum.

Thurlow, M., & Bolt, S. (2001) Empirical support for accommodations most often allowed in state policy (Synthesis Report 41). Retrieved March 4, 2004, from University of Minnesota, National Center for Educational Outcomes <http://education.umn.edu/NCEO/OnlinePubs/Synthesis41.html>

Thurlow, M., Elliott, J., & Ysseldyke, J. (1999). Out of level testing: Pros and cons (NCEO Policy Directions). Retrieved March 4, 2004 from University of Minnesota, National Center for Educational Outcomes <http://education.umn.edu/NCEO/OnlinePubs/>

Thurlow, M. Elliott, J., Ysseldyke, J. (2003). *Testing Students with Disabilities*. Thousand Oaks: Corwin Press, Inc.

Thurlow, M., & Minnema, J. (2001). States out-of-level testing policies (out of level testing report 4) Retrieved March 4, 2004 from University of Minnesota, National Center for Educational Outcomes <http://education.umn.edu/NCEO/OnlinePubs/>

- Thurlow, M., Wiley, H.I., Bielinski, J. (2003, April). Going public: What 2000-2001 reports tell us about the performance of students with disabilities. Retrieved March 18, 2004, from University of Minnesota, National Center for Educational Outcomes http://education.umn.edu/NCEO/OnlinePubs/NCLB_2001_disabilities.htm
- Turner, M., Baldwin, L., Kleinert, H., & Kearns, J. (2001). An examination of the concurrent validity of Kentucky's alternate assessment system. *The Journal of Special Education*, 34, 69-76.
- U.S. Department of Education. (2002). The No Child Left Behind Act of 2001: Executive summary. Retrieved February 11, 2004, from http://www.ed.gov/NCLB_2001/overview/intro/execsumm.html
- U.S. Department of Education. (2003). No Child Left Behind. Retrieved February 11, 2004, from <http://NoChildLeftBehind.gov/>
- U.S. Department of Education. (2003). No Child Left Behind provision gives schools new flexibility and ensures accountability for students with disabilities-Fact sheet. Retrieved March 25, 2004 from http://www.ed.gov/NCLB_2001/freedom/local/specedfactsheet.html?exp=0
- U.S. Department of Education. (2003). Title I--Improving the Academic Achievement of the Disadvantaged. 34 CFR Part 200. Retrieved March 25, 2004, from <http://www.ed.gov/legislation/FedRegister/finrule/2003-4/>

U.S. Department of Education, Office for Civil Rights. (2000). *The use of tests as part of high-stakes decision-taking for students goal in a resource guide for educators and policymakers*. Washington, DC: U.S.D.O.E.

Ysseldyke, J., & Olsen, K. (1997). Putting alternate assessment into practice: what to measure and possible sources of data. Retrieved March 4, 2004 from University of Minnesota, National Center for Educational Outcomes.
<http://education.umn.edu/NCEO/OnlinePubs/>

APPENDIX A

Perceptions of Accountability in Special Education

| | | | | |
|-------------------|----|--|----|--------------|
| | | | | 121 |
| Strongly Disagree | gr | | gr | Strongly Agr |

2. I believe the inclusion of students with disabilities in the NJ Assessment system has improved the general education teachers' curricular knowledge and ability to modify the

| | | | | |
|-------------------|---------|---------|------|--------------|
| Strongly Disagree | gr | | gr | Strongly Agr |
| 1 | 2 | 3 | 4 | 5 |
| | Disa ee | Neutral | A ee | ee |

3. In my view the inclusion of students with disabilities in the NJ Assessment system has improved general education teachers' knowledge of the general education curriculum.

| | | | | |
|-------------------|---------|---------|------|--------------|
| Strongly Disagree | gr | | gr | Strongly Agr |
| 1 | 2 | 3 | 4 | 5 |
| | Disa ee | Neutral | A ee | |

| | | | | |
|-------------------|---------|---------|------|--------------|
| Strongly Disagree | gr | | gr | Strongly Agr |
| | Disa ee | Neutral | A ee | ee |

5. In my view the curricular accommodations are consistently used for instruction and testing.

| | | | | |
|----|---------|---------|------|----|
| 1 | 2 | 3 | 4 | 5 |
| ee | Disa ee | Neutral | A ee | ee |

6. I believe the inclusion of the general assessment test scores for students with disabilities distorts my district's AYP results.

| | | | | |
|---|---------|---------|------|----|
| 1 | 2 | 3 | 4 | 5 |
| | Disa ee | Neutral | A ee | ee |

7. In my view there are more positive changes occurring in my district concerning the inclusion of all students than negative changes, including those with the most significant cognitive disabilities, in our districts assessments stem.

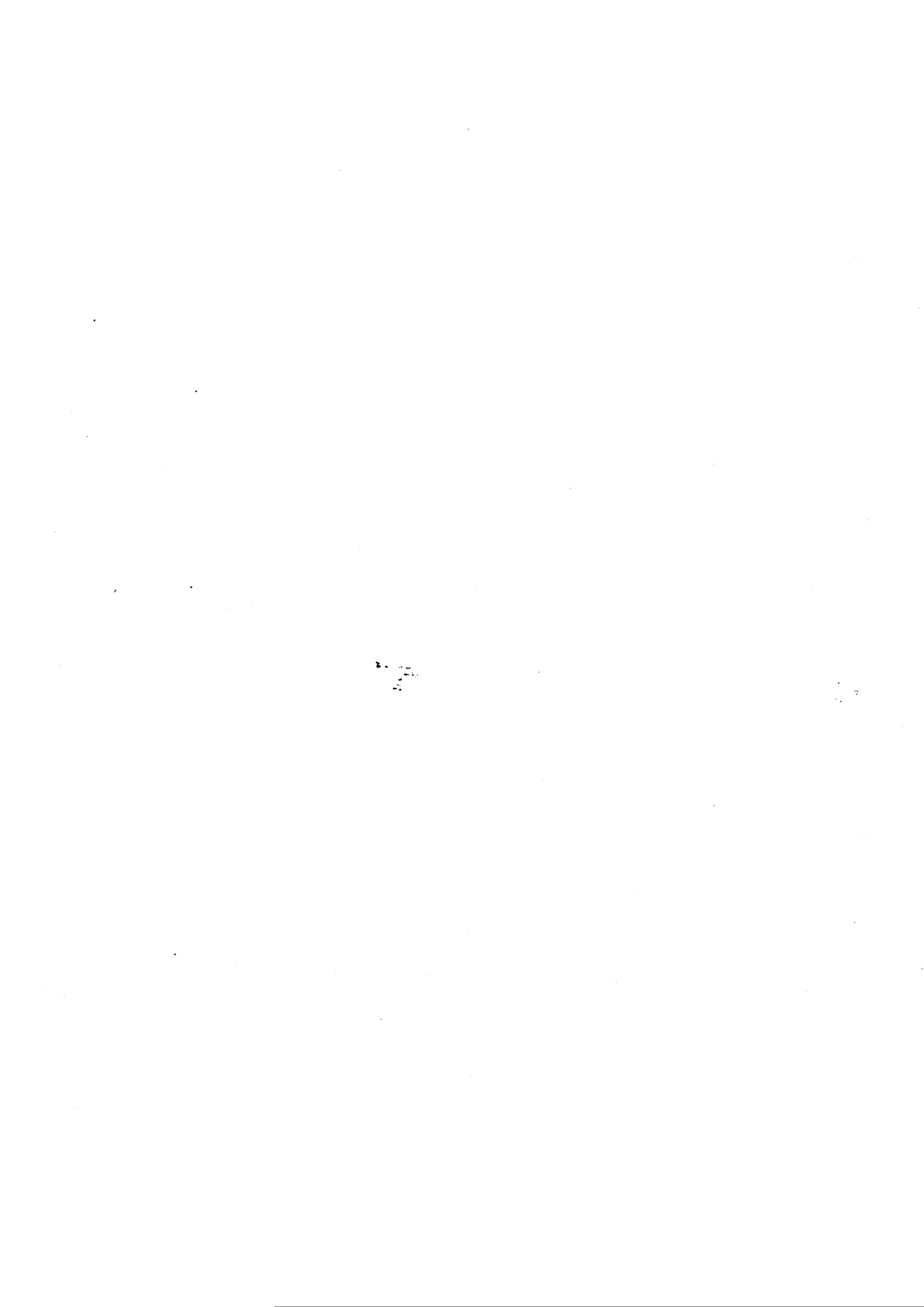
| | | | | |
|---|---------|---------|------|----|
| 1 | 2 | 3 | 4 | 5 |
| | Disa ee | Neutral | A ee | ee |

8. In my view the inclusion of the Alternate Proficiency Assessment test scores for students with disabilities distorts my district's AYP results.

| | | | | |
|---|---------|---------|------|----|
| 1 | 2 | 3 | 4 | 5 |
| | Disa ee | Neutral | A ee | ee |

9. I believe the scores for students with disabilities on the general education assessment accurately reflect student progress toward the CCCS.

| | | | | |
|---|---------|---------|------|----|
| 1 | 2 | 3 | 4 | 5 |
| | Disa ee | Neutral | A ee | ee |



SCRIPT FOR FOCUS GROUP SESSION**A Study of New Jersey's Assessment and Accountability System for Students with Disabilities and Its Impact on Special Education Services****Julie Gardner**

One of the most recent reform efforts to have a major impact on students with disabilities had been the No Child Left Behind Act and its mandate that all students participate in the state assessment process. New Jersey's assessment process is the NJ ASK 3-4, GEPA, HSPA and APA.

1. What evidence of change have you observed and experienced since the implementation of the New Jersey Assessment system relative to students with disabilities.

Probes:

- a. Resources, class size
- b. Instructional delivery options for students with disabilities
- c. Participation rates in NJ Assessment (increase or decrease)
- d. Curricular reform

2. What do you see as a positive result of the participation of students with disabilities in the assessment and accountability system.

Probes:

- a. Improved performance of students with disabilities on state/assessments
- b. Increased academic expectations for student with disabilities
- c. Increased networking between general and special educators
- d. Increased participation of special educators in training on standards and assessment
- e. Increased use of appropriate accommodation
- f. More students with disabilities accessing state academic content standards
- g. More students with disabilities in statewide general assessments

3. What do you see as a negative consequence of the participation of students with disabilities in the assessment and accountability system.

Probes:

- a. Few students with disabilities able to achieve "proficient" level on state general assessments/alternate proficiency assessment
- b. Including students with disabilities in the accountability system makes school and/or districts look less effective
- c. Too many students are identified to take the alternate assessment

An intended primary outcome of standards-based reform is improved student learning through the establishment of high standards for all students and assessment of student progress toward meeting these standards.

4. How does the NJ Assessment system do this for students with disabilities?
Probes:
 - a. In relation to the NJ Core Curriculum Content Standards
5. What impact has the NJ Assessment program and the inclusion of students with disabilities had on child study teams and teachers in your districts?
Probes:
 - a. Professional development opportunities
 - b. Knowledge of subject matter, curriculum, standards
 - c. Attitude/moral
6. What impact, if any, has the NJ Assessment program and the inclusion of students with disabilities had on the students.
Probes:
 - a. Retention and promotion
 - b. Attitude/moral
 - c. Knowledge of subject matter
7. What have the parents of your student said about the NJ assessment system
Probes:
 - a. Impact on child's learning
 - b. Parent involvement
8. What impact has the inclusion of students with disabilities in the NJ Assessment process had on you district's accountability status?
Probes:
 - a. General education assessment results
 - b. APA results
9. What approach does your district use to determine which students participate in the alternate assessment?
10. How are accommodations determined in your district?
Probes:
 - a. Accommodations used in day to day instruction and testing
 - b. Data collection on the accommodations that students use on state assessments
11. What are some of the major obstacles to including students with disabilities in the district accountability system?

13. Describe any other methods of assessment you believe would meet the assessment requirements.

Probes:

- a. IEP Review
- b. Checklist

Describe emerging issues in relation to assessment or accountability in your district or within the state.