Identifying differences in testing attitudes, practices, and concerns of teachers in outcome- and nonoutcome-based elementary schools

Janet F. Rope
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Curriculum and Instruction Commons, Educational Administration and Supervision Commons, Educational Assessment, Evaluation, and Research Commons, and the Elementary Education and Teaching Commons

Recommended Citation
Rope, Janet F., "Identifying differences in testing attitudes, practices, and concerns of teachers in outcome- and nonoutcome-based elementary schools" (1994). Retrospective Theses and Dissertations. 11315.
https://lib.dr.iastate.edu/rtd/11315

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Identifying differences in testing attitudes, practices, and concerns of teachers in outcome- and nonoutcome-based elementary schools

Rope, Janet Florence, Ph.D.

Iowa State University, 1994
Identifying differences in testing attitudes, 
practices, and concerns of teachers 
in outcome- and nonoutcome-based elementary schools 

by 

Janet F. Rope 

A Dissertation Submitted to the 
Graduate Faculty in Partial Fulfillment of the 
Requirements for the Degree of 
DOCTOR OF PHILOSOPHY 

Department: Professional Studies in Education 
Major: Education (Educational Administration) 

Approved: 
Signature was redacted for privacy. 
In Charge of Major Work 
Signature was redacted for privacy. 
For the Department and Education Major 
Signature was redacted for privacy. 
For the Graduate College 

Iowa State University 
Ames, Iowa 
1994
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Statement of the Problem</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Purposes of the Study</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Objectives of the Study</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Hypotheses to be Tested</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Basic Assumptions</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Definition of Terms</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Delimitations</td>
<td>9</td>
</tr>
<tr>
<td>II</td>
<td>Review of the Literature</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Foundational Research</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Learning as a function of time</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Learning quality, rate, and expectations</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Systemic change</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>A Nation at Risk</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>National goals</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>OBE Model</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Outcome-based Education Controversy</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Classroom Assessment</td>
<td>27</td>
</tr>
<tr>
<td>III</td>
<td>Methods</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Questionnaire Identification and Revision</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Methodology and Procedures</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Sample Selection and Data Collection</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Treatment of Data</td>
<td>37</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1.</td>
<td>Related research</td>
<td>32</td>
</tr>
<tr>
<td>2.</td>
<td>Number and percentage of distributed and completed surveys by school and type</td>
<td>35</td>
</tr>
<tr>
<td>3.</td>
<td>Demographics of respondents</td>
<td>41</td>
</tr>
<tr>
<td>4.</td>
<td>Importance of test information when diagnosing individual students' needs</td>
<td>44</td>
</tr>
<tr>
<td>5.</td>
<td>Importance of test information when diagnosing group needs</td>
<td>45</td>
</tr>
<tr>
<td>6.</td>
<td>Importance of test information when grouping students for instruction</td>
<td>46</td>
</tr>
<tr>
<td>7.</td>
<td>Importance of test information when determining achievement potential of students</td>
<td>46</td>
</tr>
<tr>
<td>8.</td>
<td>Importance of test results when assigning grades</td>
<td>47</td>
</tr>
<tr>
<td>9.</td>
<td>Importance of test information when evaluating an instructional unit</td>
<td>48</td>
</tr>
<tr>
<td>10.</td>
<td>Importance of test information to communicate academic expectations</td>
<td>48</td>
</tr>
<tr>
<td>11.</td>
<td>Importance of tests to control and motivate students</td>
<td>49</td>
</tr>
<tr>
<td>12.</td>
<td>Teachers' level of use of objective teacher-made paper and pencil tests</td>
<td>50</td>
</tr>
<tr>
<td>13.</td>
<td>Teachers' level of use of text embedded tests</td>
<td>51</td>
</tr>
<tr>
<td>14.</td>
<td>Teachers' level of use of standardized achievement tests</td>
<td>52</td>
</tr>
<tr>
<td>15.</td>
<td>Teachers' level of use of oral questioning</td>
<td>52</td>
</tr>
<tr>
<td>16.</td>
<td>Teachers' level of use of structured performance assessment</td>
<td>53</td>
</tr>
<tr>
<td>17.</td>
<td>Teachers' level of use of spontaneous performance assessment</td>
<td>54</td>
</tr>
</tbody>
</table>
Table 18. Teachers' concerns about their use of objective teacher-made paper and pencil tests 55

Table 19. Teachers' concerns about their use of text embedded tests 56

Table 20. Teachers' concerns about their use of standardized achievement tests 57

Table 21. Teachers' concerns about their use of oral questioning 57

Table 22. Teachers' concerns about their use of structured performance assessment 58

Table 23. Teachers' concerns about their use of spontaneous performance assessment 59

Table 24. Importance of tests for communicating expectations in reading and mathematics 60

Table 25. Use of standardized achievement tests by age categories 61

Table 26. Use of structured performance assessment by education level 62
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>The mastery learning instructional process</td>
<td>14</td>
</tr>
<tr>
<td>Figure 2</td>
<td>The relationship of aptitude and achievement under uniform instruction</td>
<td>15</td>
</tr>
<tr>
<td>Figure 3</td>
<td>The relationship of aptitude to achievement under optimal instruction</td>
<td>16</td>
</tr>
</tbody>
</table>
CHAPTER I. INTRODUCTION

During the 1980s and 1990s, three waves of educational reform can be delineated. The first wave began at the school and education agency level with efforts to raise standards, increase accountability, lengthen school days and years, and generally raise the rigor of American public education (Manatt, 1993; Michaels, 1988; Pipho, 1992). In 1986, the Carnegie Report evaluated the success of initial reform efforts by saying:

We are doing better on the old goals, often at the expense of making progress on the goals that count the most. Because we have defined the problem of schools in terms of decline from earlier standards, we have unwittingly chosen to face backwards when it is essential that we face forward. (pp. 15, 20)

The second-wave of educational reform began at the state level when the focus shifted to teachers and instruction (Manatt, 1993; Michaels, 1988). During the second wave, attempts were made to improve teacher performance by installing teacher performance evaluation systems, increasing salaries, and a variety of pay for performance systems.

The current third wave of school reform involves a major philosophical shift and was prompted by the National Governors Conference, the Business Roundtable, and the White House. The path for reform was to begin with setting specific goals or outcomes for schools, empowering schools to accomplish the outcomes using traditional or nontraditional methods, and then holding them accountable for documented results of improved student learning. Numerous states are developing learner outcomes and assessment strategies that focus on learning as a process, rather than content. Forty-two states have mandated movement toward
defining learner outcomes through action by either the legislature, state school board, state department, or governor initiative (Varnon & King, 1993). In a comprehensive study of school restructuring efforts, all regionally-accredited high schools in the United States were surveyed to determine the magnitude of change efforts at the high school level (Cawelti, 1994). Of the more than 3,300 responding schools, approximately a fourth reported that outcome-based education to be in general use, with another third having partially implemented this approach, and 14.1% reporting that they planned to introduce the change in the next year. Slightly more than a fourth had no plans to implement this innovation during the coming year.

A major obstacle to the success of school restructuring is that it will be measured against the old model of basic skills achievement with the old measures of norm-referenced, machine-scorable tests (Lewis, 1992; Tucker, 1988). Reform begins by recognizing that testing is central to instruction. The creation of student assessment alternatives to standardized multiple-choice tests is an essential feature of redesigned schools (Herman, Aschbacher, & Winters, 1992; Lewis, 1992; O'Neil, 1990; Wiggins, 1992). When an educational problem persists despite the well-intentioned efforts of many people to solve it, it's a safe bet that the problem hasn't been properly framed (Wiggins, 1992). Assessment in education has clearly become such a problem according to Wiggins. Standardized tests are based on behavioral research from 30 years ago and are incongruent with current research; they are skill, not process, oriented; they are used to sort and classify youngsters rather than to
give direction for learning; they are incomplete; and they have a mistaken aura of objectivity (Heald-Taylor, 1989).

A testing program must address questions about the inevitable impact of tests (and scoring methods) on students and their learning (Wiggins, 1989). Because testing practices set the *de facto* standards of a school, reform begins, then, by recognizing that the tests are central to instruction. They not only monitor standards, but also set them (Wiggins, 1989).

**Statement of the Problem**

The primary problem upon which this study focuses is whether implementation of the outcome-based education (OBE) model is accompanied by a change in teacher attitudes and behaviors concerning student assessment. The problem is to answer the following questions: 1) Are teachers' attitudes about the importance of assessment-related decisions different in outcome-based and nonoutcome-based schools? 2) Are teachers' levels of use of various assessment types different in the two types of schools? and 3) Are teachers' concerns about the use of a variety of assessment types dependent on affiliation with either an outcome-based or nonoutcome-based school? It is essential that efforts at reform be based on ideas that are important to those who must carry them out. The belief systems of the people who make the decisions and do the work need to be altered (Aspy, Aspy, & Quinby, 1993; Combs, 1988).
Purposes of the Study

The purpose of this study was to determine whether the transformational changes that OBE was designed to produce were realized. Was OBE really working in the schools where it had been selected as the model for restructuring? The study probed assessment practices in a sample of elementary school teachers selected from three outcome-based and one nonoutcome-based school district. The primary purpose of this study is to find whether implementation of the outcome-based school model is evidenced by a shift in teacher attitudes, practices, and concerns about student assessment. As a result of the review of literature, an ERIC abstract describing a related study was found that evaluated teacher attitudes, concerns, and practices toward several types of assessment. Efforts to locate the author eventually resulted in the identification of the specific instrument used, Teacher's Self-analysis of Classroom Assessment Procedures, and instrument author, Richard J. Stiggins, Director, Center for Classroom Assessment, Northwest Regional Laboratory, 101 S.W. Main, Suite 500, Portland, Oregon 97204. During telephonic contact, Dr. Stiggins granted permission for the instrument to be adapted and used for the proposed research in outcome- and nonoutcome-based elementary schools. Stiggins also agreed to send copies of findings from his research using the instrument. The original study attempted to determine the current state of teachers' assessment attitudes and practices in order to design professional development training to improve student assessment practices. With minor deletions and revisions, it was concluded that the instrument was suitable for the purposes of this study.
Objectives of the Study

In order to accomplish the purposes of this study, the following tasks were undertaken:

1. Identify and analyze the essential elements of OBE.
2. Determine whether the transformational changes attributed to OBE are accomplished during implementation.
3. Identify a survey instrument through a review of the literature that probes teacher attitudes concerning student assessment.
4. Adapt the survey instrument to meet the specific requirements of the study.
5. Administer the survey to a sample of elementary school teachers in grades one through five in outcome-based and nonoutcome-based school districts.
6. Analyze the results of the survey to determine whether teachers' attitudes about assessment related decisions, levels of use of assessment types, and concerns about assessment types vary by the type of school.
7. Develop conclusions based upon the findings of the study regarding testing attitudes, practices, and concerns.

To reach the objective of this study, the questionnaire was identified, adapted, field tested and revised, and administered to elementary teachers in grades one through five, regarding student assessment. A five-point Likert-type scale was used to rate the importance of various types of classroom assessment related decisions.
The scale to rate the importance of classroom decisions ranged from important to unimportant.

To determine the extent to which each assessment option was used and the types of concerns teachers had about their use, scaling systems were adapted from ones developed by the University of Texas Research and Development Center in Teacher Education.

A four-point scale was employed to determine the level of use for each of the following six types of assessment instruments: teacher-made tests; text embedded tests; standardized achievement tests; oral question and answer; structured performance assessment; and spontaneous performance assessment. The level of use scale ranged from nonuse to regular use.

A seven-point scale was used to ascertain the primary concern teachers report about the use of each of the six types of assessment. The scale ranged from concern about a lack of information to no concerns about the assessment type.

Hypotheses to be Tested

This investigation sought to use teachers' attitudes, practices, and concerns as a measure of change to provide insight as to whether implementation of OBE resulted in fundamental change in the structure of schooling. Specific null hypotheses tested were:

1. Teachers' attitudes concerning the relative importance of different types of educational assessment decisions are independent of whether the elementary school teacher is in an outcome- or nonoutcome-based school district in Iowa.
2. Assessment practices are independent of whether the elementary school teacher is in an outcome- or nonoutcome-based school district in Iowa.

3. Assessment concerns are independent of whether the elementary school teacher is in an outcome- or nonoutcome-based school district in Iowa.

Basic Assumptions

This study was predicated on the following assumptions:

1. That assessment attitudes, practices, and concerns can be accurately measured using a written survey.

2. That participating teachers will select answers that accurately reflect their attitudes, practices, and concerns.

3. That the definition of each assessment is clearly delineated.

4. That teachers understand and are able to differentiate between the six types of assessments.

5. That teachers are able to focus on each different type of assessment when answering questions about practices and concerns.

6. That the teachers volunteering to take part in the study will represent the attitudes, practices, and concerns found in the population.

7. That change in testing attitudes, practices, and concerns can be used as an indication of genuine educational reform.

8. That OBE is well enough established in the selected schools for differences to be accurately measured.
Definition of Terms

Words often have different meanings depending on their context. The following definitions will be used in this investigation.

1. **ASSESSMENT**: The act of collecting and analyzing data in order to determine the quality of student performance.

2. **ATTITUDES**: The cognitive and affective variables describing a person's mental set toward another person, thing, or state.

3. **CONCERNS**: A matter that produces anxiety, apprehension, or worry.

4. **NONOUTCOME-BASED DISTRICT**: For the purpose of this study, this is a district that describes itself as a traditional district and is not currently involved in a systematic restructuring effort.

5. **OBJECTIVE TEACHER-MADE PAPER AND PENCIL TESTS**: This assessment category includes all true/false, multiple choice, matching, fill-in, and short answer tests and quizzes which the teacher develops.

6. **ORAL QUESTIONING**: This type of assessment is used on a day-to-day basis during instruction to track whether individual students or the class as a group are learning the material.

7. **OUTCOME-BASED DISTRICT**: For the purpose of this study, this is a district that identifies itself as implementing the Spady model of outcome-based education and is a member of the outcome-based consortium known as the Iowa Network for Success.

8. **PRACTICES**: Habitual or customary actions or ways of doing something.
9. **SPONTANEOUS PERFORMANCE ASSESSMENT:** This type of assessment includes informal opportunities to observe and evaluate a student's performance and to judge the student's proficiency.

10. **STANDARDIZED ACHIEVEMENT TEST:** These instruments are constructed by professionals and administered with standard directions and under standard conditions.

11. **STRUCTURED PERFORMANCE ASSESSMENT:** This type of assessment includes 1) a clearly defined reason for assessment; 2) planned exercises to elicit student responses; 3) a prespecified response to be evaluated; and 4) carefully spelled out scoring procedures.

12. **TESTS EMBEDDED OR INCLUDED IN THE INSTRUCTIONAL MATERIALS:** These assessment instruments may be found in an instructor's guide or may take the form of questions at the end of chapters in the materials themselves.

**Delimitations**

Only elementary classroom teachers in grades one through five were used as subjects for this study. A total of 96 subjects from four Iowa school districts were included in the study. Three of the districts were classified as outcome-based districts and one was a nonoutcome-based district. Subjects were asked to focus on either mathematics or reading when responding to questionnaire items.

The Iowa State University Committee on the Use of Human Subjects in Research insists that researchers make sure that the rights and welfare of the human subjects are adequately protected, that risks are outweighed by
the potential benefits and expected value of the knowledge sought, that confidentiality of data is assured, and that informed consent be obtained by appropriate procedures. These procedures were approved by the Committee and closely followed in this study. Consent to participate in the project in the form of modified consent, was assumed by those voluntarily completing and returning the questionnaire.
CHAPTER II. REVIEW OF THE LITERATURE

The review of literature for this study focused on the broad historical research foundation that contributed to the current outcome-based education (OBE) movement, the contemporary philosophy and key operational principles espoused by advocates of OBE, controversy surrounding the OBE movement, and the broad area of student assessment. This review of literature was based on the premise that change in student assessment accompanies general education reform and therefore can be viewed as an indicator of change.

The literature was used to identify ideas and events in American education during the past thirty years that created the foundation for outcome-based education, delineate the major principles of the OBE model for educational change, characterize the controversy concerning OBE, and create a current view of student assessment in the United States.

Foundational Research

Several events in American education occurred during the past thirty years that contributed to an atmosphere that was ripe for a change model such as OBE.

Learning as a function of time

In 1963, John Carroll published his landmark article, "A Model for School Learning." In his article, he challenged the idea that the purpose of education was to sort and select students for future work as a result
of a demonstrated ability to learn. He contended that the level or degree of learning was largely a function of the relationship between the amount of time allowed for learning, and the amount of time needed by the learner. The revolutionary idea that many students, given enough time, could learn at levels that had previously been reserved for only the mentally strongest, captured the attention of many in the education world.

Quality of instruction and other variables were also addressed by Carroll, but the concept of time as a function of level of learning, in particular, offered educators hope in their attempts to successfully educate all students. Carroll's theoretical work laid the foundation for practical application by Bloom and other educational researchers.

Learning quality, rate, and expectations

Three major instructional constructs were identified by Benjamin Bloom (1981), which resulted in the creation of Learning for Mastery. Bloom viewed those constructs as controlling factors in the way schools were organized. Substantive educational change would depend on a shift in the belief system concerning each of those instructional constructs.

The first construct can be described as the poor student/good student concept. It supported the select and sort philosophy described earlier. The poor student/good student idea is reflected in the belief that some students have "it" and others don't. It provided justification for tracking students into high and low level classes, effectively dooming students in low classes to a "watered down" education.
Norm-referenced tests are instruments that provide the primary support for the poor student/good student concept. The method for creating tests was described by Smith (1992):

These tests are methodically and carefully designed to create variance—a distribution of scores off a mean score to determine how students compare to each other. Tests that create the variance are deemed "sensitive." They are not considered sensitive if large numbers of student scores cluster around the mean score. If that phenomenon occurs, the psychometricians return to the drawing board to construct other questions which will generate the variance. Without the variance, students cannot be so easily kenneled into one track or another. Yet the variance is artificially generated. (p. 6)

The effect of norm-referenced tests on students seemed to confirm their membership in either the poor or good student group, often inhibiting future learning.

The second major construct identified by Bloom can be referred to as the fast learner/slow learner concept. While the first construct supported the sort and select method to identify and weed out the poor learner, the fast learner/slow learner concept offered a method to accomplish it. The implication of the second construct was that while some students were considered good learners because of the speed with which they learned, some were considered flawed because of the amount of time needed to learn. The rate at which students learn was considered more important than the potential level of achievement.

The third construct included speculation by Bloom that all students could achieve at high levels, and learn at high rates and at high levels of motivation if favorable learning conditions were in place. This construct was closely associated with Learning for Mastery. Qualitative
research about mastery learning supports the idea that all students can learn (Andersen, 1973; Block, 1974; Guskey, 1985).

Bloom (1981) believed that dividing material into small sequential units and assessing students' learning at the end of each unit was a useful instructional technique. The instructional model was further enhanced by including meaningful feedback to students for grading and diagnostic purposes, and corrective processes for remediation or reteaching. Figure 1 illustrates the mastery learning process.

Figure 1. The mastery learning instructional process (adapted from Guskey, 1988)

A meta-analysis of findings from 108 controlled evaluations showed that mastery learning programs have positive effects on the examination performance of students in colleges, high schools, and the upper grades in elementary schools (Kulik, Kulik, & Bangert-Drowns, 1990). Figure 2 illustrates the relation of aptitude and achievement under uniform instruction.
When instruction was uniform, with tests and quizzes given as a summative evaluation, the amount of achievement is fairly dependent on student aptitude. Bloom discovered that in most cases only about 20 to 30 percent of the students in a class really learn well what the teacher sets out to teach (Cusky, 1988). As a result of using mastery learning, Bloom believed that the system of formative testing and systematic correction of learning problems would provide students with an "optimal instruction" resulting in higher levels of achievement. Figure 3 represents the relationship of aptitude to achievement under "optimal instruction."

Mastery programs have positive effects on student attitudes toward course content and instruction but may increase student time on instructional tasks. While Slavin (1990) agreed with the positive effect mastery learning had in terms of focusing teachers on a given set of objectives, he questioned the claim that mastery learning could accelerate achievement in elementary and secondary schools.
An important contribution to the foundation of outcome-based education came from John Champlin (1986), which helped propel the movement forward. He recognized that "quick fix" solutions stood little chance of success. While he recognized it was necessary to recreate curriculum and instruction, he realized it was not sufficient to transform schooling. The entire organization had to be restructured to create a climate which would allow administrators, teachers, and everyone else connected with students to transform education. Champlin has influenced the restructuring movement with his insistence on the need for systemic change (Smith, 1992).

A comprehensive management plan for changing schools must have certain structural parts regardless of the size of the district or the change planned (Wiles, 1993).
• It must identify a reality (needs assessment).
• It must have goals and priorities (policies and standards).
• It must have organization (design and task analysis).
• It must be logical (management and implementation plan.)
• It must be seen as successful (evaluation or validation of results).

Wiles (1993) further cautioned that lasting change in schools must be implemented and closely monitored to ensure that it actually becomes institutionalized.

A Nation at Risk

In 1981, Secretary of Education T. H. Bell created the National Commission on Excellence in Education as a result of his concern about "the widespread public perception that something is seriously remiss in our educational system." The Commission's charter contained the following charges (1983):

1. assessing the quality of teaching and learning in our nation's public and private schools, colleges, and universities;
2. comparing American schools and colleges with those of other advanced nations;
3. studying the relationship between college admissions requirements and student achievement in high school;
4. assessing the degree to which major social and educational changes in the last quarter century have affected student achievement; and
5. defining problems which must be faced and overcome if we are successfully to pursue the course of excellence in education.
The investigation by the Commission resulted in the publication of A Nation at Risk (1983). The premise upon which the Commission undertook its work, molded and guided the inquiry.

All, regardless of race or class or economic status, are entitled to a fair chance and to the tools for developing their individual powers of mind and spirit to the utmost. This promise means that all children by virtue of their own efforts, competently guided, can hope to attain the mature and informed judgment needed to secure gainful employment, and to manage their own lives, thereby serving not only their own interests but also the progress of society itself. (p. 4)

A Nation at Risk effectively signaled the public that the premise guiding its work, a fair chance for educational growth for all, was in danger of being lost. The report detailed findings and recommendations in five areas: 1) content, 2) expectations, 3) time, 4) teaching, and 5) leadership and fiscal support.

The content was defined as the "stuff" of education, the curriculum. Findings reported that the curriculum was homogenized, diluted, and diffused to the point that secondary schools no longer had a central purpose. The curriculum was described as cafeteria-style, in which the appetizers and desserts were easily mistaken for the main courses. It was recommended that state and local high school graduation requirements be strengthened and that, as a minimum, all students seeking a diploma be required to lay the foundations in the five new basics by taking four years of English; three years of mathematics; three years of science; three years of social studies; and one-half year of computer science. Two years of foreign language in high school were strongly recommended for college bound students.
Expectations were defined in terms of the level of knowledge, abilities, and skills school and college graduates should possess. The most significant expectation deficiencies were: a decrease in the amount of homework and achievement while grades improved; the small amount of time, relative to other industrialized nations, spent in mathematics, biology, chemistry, physics, and geography; and the low standards for entrance to public colleges. Recommendations called for more rigorous and measurable standards, and higher expectations for academic performance and student conduct. Colleges and universities were encouraged to raise their requirements for admission.

Three disturbing facts about American schools and students' use of time were: 1) Compared to other nations, American students spent much less time on school work; 2) time spent in the classroom and on homework was often used ineffectively; and 3) schools were not helping students develop the study skills to use time well or the willingness to spend more time on school work. Commission recommendations included a longer school day or a lengthened school year, and a more effective use of the time spent in school.

Major findings concerning the quality of teaching reported the following deficiencies: too little teacher preparation in subject matter courses; low teacher salaries; and the lack of qualified teachers, especially in mathematics and science. Recommendations included the redesign of teacher preparation to require high educational standards. Teacher salaries were to be increased and tied to an effective evaluation system with peer review. An 11-month contract for teachers was
recommened to ensure time for curriculum and professional development and a more adequate level of teacher compensation. The need to encourage outstanding students to enter the teaching profession through the use of incentives was also recognized. The Commission also promoted the involvement of master teachers in the design of teacher preparation programs and in supervising probationary teachers.

It was further recommended that the public hold educators and elected officials responsible for providing the leadership necessary to achieve the recommended reforms, and that the fiscal support and stability required to bring about the reforms be provided.

Many states responded to *A Nation at Risk* with legislative action establishing mandates, accountability directives, and various other changes in education policies. Many states created their own commissions to study their educational systems and recommend reform measures (Bell, 1993; Crosby, 1993). The report also spawned a new generation of reports on teacher training, youth employment, higher education, vocational education, science, math, and schooling in general (Crosby, 1993).

Perhaps Bell best summed up the effects of *A Nation at Risk* ten years after its publication.

The 10 years since the publication of *A Nation at Risk* have been a splendid misery for American education. We have learned much. We have suffered many disappointments. But we have not given up the quest to shape education into the super-efficient enterprise that it must become if America is to keep its proud place of leadership in the marvelous Information Age of this decade and beyond. Perhaps we should have made much more progress than we have. But at least we have stayed with the task. (p. 593)
The charges brought by the publication of *A Nation at Risk* paved the way for the outcome-based education restructuring model.

**National goals**

In 1989, President Bush convened the Education Summit with the nation's governors, and embarked on an historic venture to change the national educational emphasis from process to performance, from complacency to high expectations. In response to the Summit, the National Education Goals Panel was created. In 1994, President Clinton signed into law the Goals 2000: Educate America Act. The law ties federal dollars to the following eight goals and to the "school delivery standards" that are to measure the school conditions necessary to allow students to meet the academic standards:

1. **Goal 1.** All children in America will arrive at school ready to learn.
2. **Goal 2.** The high school graduation rate will increase to at least 90 percent.
3. **Goal 3.** Students will master challenging subject matter.
4. **Goal 4.** Teachers will have access to training programs to improve their skills.
5. **Goal 5.** By the year 2000, U.S. students will be first in the world in math and science.
6. **Goal 6.** All adult Americans will be literate and able to compete in a global economy.
7. **Goal 7.** Every school in America will be free of drugs and violence.
8. **Goal 8.** Every school will strive to increase parental involvement and participation in their children's education.

Numerous groups are developing goals and standards. Goals and standards for mathematics, science, geography, and civics are currently available. During 1993-94, the U.S. Department of Education is funding standards projects in the following areas: the arts, civics, English,
foreign language, history, mathematics, and science. Each report will include goals, standards of performance, and suggested assessment techniques with a year 2000 target date for implementation.

OBE Model

The origins of outcome-based education lie in the theory of mastery learning as developed by John B. Carroll (1963) and the extension of this theory and the development of practical implications as a result of research by Bloom (1968, 1976). Several educational researchers including Block (1971), Anderson (1975), and Guskey (1985) have further elaborated upon the work of Carroll and Bloom.

During the early 1980s, the ideas and practice of mastery learning grew from the classroom to the larger arena of the total school program, which necessitated a more comprehensive articulation of the philosophical premises and instructional components of this theory about student learning (Desmond, 1992). Advocates and preeminent practitioners of mastery learning formed the National Center for Outcome-Based Education and the Network of Outcome-Based Schools as a means of developing a unified statement of the essentials of mastery learning for the classroom, school, and school district (Block, Efthim, & Burns, 1989).

William G. Spady, who has had a distinguished career including basic research with the Far West Laboratory at San Francisco and as an associate executive director of the American Association of School Administrators, is generally credited with formally structuring and popularizing the theoretical framework for outcome-based education. Spady insists that OBE
is extremely elementary in concept. The Spady model of OBE is founded on the beliefs that 1) all students can learn and succeed, 2) success breeds success, and 3) schools control the conditions of success. According to Spady (1988):

Outcome-based education (OBE) means organizing for results: basing what we do instructionally on the outcomes we want to achieve, whether in specific parts of the curriculum or in the schooling process as a whole. Outcome-based practitioners start by determining the knowledge, competencies, and qualities they want students to be able to demonstrate when they finish school and face the challenges and opportunities of the adult world. Then, with these “exit outcomes” clearly in mind, they deliberately design curriculums and instructional systems with the intent that all students will ultimately be able to demonstrate them successfully. OBE, therefore, is not a “program” but a way of designing, developing, delivering, and documenting instruction in terms of its intended goals and outcomes. (pp. 4, 5)

Operationally, OBE means 1) using clearly defined outcomes for all students to define and develop curriculum and establish measures of student and program success; 2) organizing instructional delivery based on the performance capabilities and learning needs of students; 3) adjusting instructional time and learning opportunities to enable all students to reach outcome goals successfully; 4) formally acknowledging and documenting student learning and success whenever they occur; and 5) modifying the instructional program on the basis of documented student learning results and available data on instructional effectiveness (Spady, 1988).

Outcome-based Education Controversy

The statewide movement toward outcome-based education began with the Iowa Department of Education and its director, Dr. William Lepley. More
than 200 Iowa educators, under the leadership of Lepley, worked for two years developing a new educational model for the state with the following nine student outcomes identified:

- Lifelong Learning,
- Problem Solving,
- Communication,
- Group Membership,
- Commitment to Quality,
- Creativity,
- Diversity,
- Environmental Responsibility, and
- Life Management.

Critics described the plan as another state conspiracy to infuse social and political values into the curricula. Some said that the state would eventually test children on values and give remedial lessons to those who failed (Siebert, March 22, 1993). "If people will be patient, we'll put something together that Iowans will be excited about and confident with," stated Lepley in an attempt to defend the goals.

The religious right was swift and effective in its counterattack. Pamphlets appeared in religious bookstores alerting the faithful of the danger of OBE and urging them to protest at state hearings. Parents for Traditional Choices groups were created to ask for basic skills instead of OBE. "Stop Outcome-Based Education" bumper stickers were produced and sold well (Manatt & Dripps, 1994). A Davenport school board member sued
the superintendent to see the test that would be used to measure a student’s progress toward the outcomes.

Despite a counterattack in early April by the Iowa State Education Association, the controversy over OBE continued. Plans to reconsider the nine outcomes were developed as a result of the mixed reviews. The review process was to begin on March 1, public review in May, and approved by the State Board in August (Knowles, January 1993).

On May 6, the Iowa Department of Education abandoned its plan to establish statewide student performance outcomes. "I don’t have enough support in the state to move it forward," Lepley said. While he said that the department was not backing away from the outcomes philosophy, instead of mandating statewide goals, the department would help school districts establish their own outcomes at the local level.

The Iowa experience with OBE is not an isolated example. OBE proponents in Kentucky, Oklahoma, Connecticut, Montana, Tennessee, Michigan, New Mexico, Arkansas, Virginia, Minnesota, and Chicago have experienced similar attacks, frequently led by religious conservatives (Pipho, 1994).

The spotlight was on the Kentucky Education Reform Act when Robert Saxton, director of the Prichard Committee, released materials being used by the forces opposed to the reforms. He detailed the following eight lines of argument used by critics of OBE:

• Schools are teaching satanism, homosexuality, socialism/communism, and anti-American and anti-Christian views.
• Reform is "leveling down or pulling down" all children to the lowest level.
• Kentucky teachers and schools are "indoctrinating children with social values" instead of emphasizing academic skills and knowledge.
• The growing presence of technology and computers is not for instructional purposes but for building dossiers on children and parents.
• Reforms such as Kentucky's form part of a national conspiracy to promote "highly centralized government control of children and their minds."
• Basic facts are most important; exploring alternatives, thinking things through for oneself, and analyzing concepts and ideas should not be promoted by the public schools.
• Local control is being supplanted by state control through accountability, a focus on outcomes, and changes in the assessment process.
• The Kentucky Education Reform Act is an untried, untested experiment foisted on state legislators by various national organizations and "educrats."

Connecticut's reform plan, led by the state's Commission on Excellence, went in the public hearing phase in early January. The commission recommended setting high standards for all students and put forth a plan for redesigning the schools. The comprehensive plan included an expanded Head Start program; upgrades in technology; an updated
curriculum to give students the skills needed to join the work force, enter higher education, or to pursue lifelong learning; varied assessment techniques to measure students' ability to demonstrate and apply knowledge; and a provision to hold teachers accountable for students' meeting performance standards.

Opposition came from the upper-class suburbs and from political and religious conservatives (Pipho, 1994). For both groups, OBE raised fears of mediocrity, mind control, and the usurping of family values.

Similar attacks took place in Oklahoma where Spady worked with several districts using OBE during pilot projects. Spady found himself defending OBE against what he considers to be outrageous and inaccurate charges. He jokingly insisted that he's not the devil reincarnate (Manatt & Dripps, 1993).

In summary of the controversy surrounding OBE, Pipho (1994) characterized the debate over OBE by stating the following:

The old rules of civility, decency, and respect for good data that used to characterize the process of debating public issues seem to have become roadkill on the information highway. The new rule appears to be "My side will look good to the extent that I can make your side look bad." Distorted information delivered in a legislative hearing or an op-ed piece, even if rebutted, will enter the public consciousness and spread as if it were fact. (p. 510)

Classroom Assessment

As long as we hold simplistic monitoring tests to be adequate models of and incentives for reaching national intellectual standards, student performance, teaching, and our thinking and discussion about assessment
will remain flaccid and uninspired (Wiggins, 1989). Central to the issue of educational reform is student assessment.

Evidence suggests that the dominant view regarding measurement in education through the early 1980s was one of documenting student achievement by using collections of standardized paper and pencil test items (Stiggins, 1992). At that time, nearly all major studies of testing in schools had focused on the role of standardized tests. The only standards on acceptable testing practices were the Standards for Educational and Psychological Tests, which only identify the ethical responsibilities of publishers of standardized paper and pencil tests (Standards, 1985). The primary source of public analysis of educational tests was the Buros Mental Measurements Yearbook series, which dealt only with published tests. A comprehensive review of research on educational testing led researchers Lazar-Morris, Polin, May, and Barry (1980) to conclude the following:

In-class assessments made by individual teachers have yet to be examined in depth. How these and other assessments are united with teacher instructional decision-making processes and how they affect classroom organization and time allocation to other objectives are areas that should be explored. Teachers place greater reliance on, and have more confidence in, the results of their own judgments of student performance, but little is known about [these] kinds of activities. (pp. 24-25)

Although most research on testing in schools has focused on attitudes toward standardized tests, a few studies provide a glimpse of attitudes about classroom assessment. The following key dimensions of classroom assessment that have been the focus of research were identified by Airasian (1984):
• Scholastic and social variables are both measured in classroom assessment.

• The relative importance of the two variables changes depending on grade level, with social factors seen as more important in elementary school.

• Teachers "size up" students as individuals, group them very quickly, and these initial estimates remain quite stable.

• Students appear sensitive to these early teacher assessments, learn their positions in the "pecking order" of the class, and respond accordingly.

• Teachers interact differently with students they perceive to be of high or low ability.

• Teachers can accurately predict student test performance and therefore use standardized test results to corroborate their own judgments.

In an attempt to document quality control of teacher-made assessments, Gullickson and Ellwein (1985) surveyed midwestern teachers as to their testing strategies. Conclusions reached by the researchers were that teachers have not been taught how to evaluate their test items, take necessary steps to improve quality, or accurately set criterion levels for student performance. Further, teachers do not value statistical analysis of test items as a helpful strategy in the classroom.

Teachers' attitudes as to the importance of various forms of assessment in making various classroom decisions were studied by Dorr-Bremme and Herman (1986). A national sample of teachers were asked to rate the importance of five types of tests when making various classroom decisions. Results led the researchers to conclude that teachers prefer their own assessments and rely most heavily on teacher observation and opinion.
Some questions concerning teacher-developed assessments were included in a study about student attitudes about testing (Beck, 1979). Results suggest that students are more concerned about teacher-made tests than standardized ones. Most students thought teacher-made tests were harder, and twice as many got nervous before a teacher-made test.

In an effort to broaden understanding of the nature and quality of teacher-developed assessment, researchers at the Northwest Regional Educational Laboratory designed a study to ascertain teachers' patterns of test use, concerns about assessment, and use of performance assessment. Two hundred and twenty-eight teachers from a range of grades, subjects, and school districts completed an extensive questionnaire describing the nature of their classroom assessment. Results allowed researchers to conclude the following:

- Teachers use their own objective tests more frequently than other assessments for all purposes.
- Teachers report heavy use of structured and spontaneous performance assessment.
- Published tests consistently play a secondary role to teacher-made tests.
- Nearly three-quarters of teachers expressed concerns about their own tests, with the most common concern having to do with improving the quality of teacher-made tests.
- While over three-quarters of teachers reported using structured performance assessments in their classroom, most stated that no purpose for the test, scoring criteria, levels of performance were identified prior to the test.

Results of the previous study motivated Stiggins and others at the Northwest Regional Laboratory to conduct further research into classroom assessment of thinking skills (1986). The research design consisted of
teacher interviews, classroom observations, analyses of written assessment documents and accompanying texts. Results for written assessments indicate a heavy reliance on recall of facts and information in science, social studies, and language arts. A striking difference was noted in mathematics, where 72 percent of all items call for inference on the part of the student. Results of oral questioning revealed that evaluation and comparison skills were largely ignored, especially in the elementary grades. Data also revealed that almost all teachers had participated in one or more workshops on teaching thinking skills, a third reported having "no training in assessing thinking skills."

Table 1 shows a summary of important research in the area of classroom assessment.


<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Date</th>
<th>Topic</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airasian</td>
<td>1994</td>
<td>Nature of classroom assessment</td>
<td>Six key dimensions were identified where research into classroom practices had been conducted.</td>
</tr>
<tr>
<td>Dorr-Bremme, Herman</td>
<td>1993</td>
<td>Use of various types of tests</td>
<td>Teacher-developed tests dominate classroom assessment.</td>
</tr>
<tr>
<td>Guillickson</td>
<td>1992</td>
<td>Quality of teacher-made tests</td>
<td>Teachers have not been taught how to evaluate their test items, take necessary steps to improve quality, or accurately set criterion levels for student performance. Teachers do not value statistical analysis of test items as a helpful strategy in the classroom.</td>
</tr>
<tr>
<td>Lazar-Morris, Polin, May, Barry</td>
<td>1980</td>
<td>Review of research on testing in schools</td>
<td>Teacher-made assessments are largely unstudied even though teachers have more confidence in and rely heavily on their own judgments of students.</td>
</tr>
<tr>
<td>Stiggins, Bridgeford</td>
<td>1985</td>
<td>Nature and quality of teacher-made assessments</td>
<td>Teachers rely heavily on teacher-made tests. Teachers' concerns about their own tests mainly concern improving the quality. While most teachers use structured performance assessments, few have predetermined purposes, scoring criteria, or performance levels.</td>
</tr>
<tr>
<td>Stiggins, Griswold, Wikelund</td>
<td>1989</td>
<td>Measuring thinking skills</td>
<td>Written assessment of thinking skills largely focuses on recall information. Oral assessment of students ignores comparison and evaluation levels of thinking.</td>
</tr>
</tbody>
</table>
CHAPTER III. METHODS

This study attempted to determine whether a difference in testing attitudes, practices, and concerns existed between elementary school teachers in outcome-based and nonoutcome-based school districts in Iowa. In addition, it is anticipated that those differences, where they exist, can service as an indication of genuine educational reform.

The identification and revision of the questionnaire, the identification of the subjects participating, procedures for data collection, and the statistical analysis are examined in this chapter.

Questionnaire Identification and Revision

While conducting a review of literature in the related areas of outcome-based education and student assessment, an abstract of a research article described an attitude survey that was used to ascertain teacher attitudes, practices, and concerns in the area of student assessment. Efforts to locate the author of the questionnaire eventually resulted in the identification of the specific instrument used, *Teacher Self-Analysis of Classroom Assessment Procedures*, and instrument author, Richard J. Stiggins, Director, Center for Classroom Assessment, Northwest Regional Laboratory. During telephonic contact, Dr. Stiggins granted permission for the instrument to be used for the proposed research in outcome-based and nonoutcome-based elementary schools.

The original questionnaire was shortened to retain only items directly related to the focus of this study. The abbreviated
questionnaire was then field tested with a sample of ten classroom teachers. Following the field test, additional revisions were made to further refine and clarify the remaining 20 items.

The instructions for completing the questionnaire asked each elementary classroom teacher to focus on one of two subject areas (reading or mathematics) and the grade level of the teacher's current assignment. Teachers were also asked to carefully distinguish between the six types of assessments as defined in the instruction section of the questionnaire. In addition, teachers were cautioned to mark answers based on actual assessment activities—not as the teacher might think they should be. Directions were supplied on every questionnaire and answers were to be marked directly on the questionnaire.

Methodology and Procedures

This study was designed to probe assessment practices in a sample of elementary school teachers from three outcome-based and one nonoutcome-based school district in Iowa using a self-analysis questionnaire. Each district was asked to administer the questionnaire to classroom teacher volunteers in grades one through five. Completed questionnaires were individually returned to the researcher in sealed envelopes. All districts responded with completed questionnaires, however, the percentage of returns differed substantially across districts. A total of 96 completed questionnaires were returned, 66 from the three outcome-based districts and 30 from the single nonoutcome-based district. Table 2 shows the number and percentage of returns from each of the four schools.
Table 2. Number and percentage of distributed and completed surveys by school and type

<table>
<thead>
<tr>
<th>School and type</th>
<th>Distributed</th>
<th>Returned</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome-based schools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School A</td>
<td>15</td>
<td>3</td>
<td>20.00</td>
</tr>
<tr>
<td>School B</td>
<td>26</td>
<td>18</td>
<td>69.23</td>
</tr>
<tr>
<td>School C</td>
<td>112</td>
<td>45</td>
<td>40.18</td>
</tr>
<tr>
<td>Total</td>
<td>153</td>
<td>66</td>
<td>43.14</td>
</tr>
<tr>
<td>Nonoutcome-based school</td>
<td>60</td>
<td>30</td>
<td>50.00</td>
</tr>
</tbody>
</table>

No explanation could be found for the low number of returns for school A.

Sample Selection and Data Collection

Initial identification of the outcome-based schools was accomplished with the assistance of the director of the Iowa Network for Success. In 1988, the Iowa Success Network was established with eight charter member school districts. The purpose of the network was to promote and share ideas concerning the implementation of outcome-based education in Iowa. The number of school districts involved in the network rose to a peak 24 and currently has 15 school district members. The network director, Al Rowe, was asked to identify the network members who had been the most successful in implementing outcome-based education, resulting in the naming of four school districts. Three of the four named schools agreed
to participate in this study. The nonoutcome-based school was identified with the assistance of Professor Manatt based on prior association with the superintendent of the district.

The resulting sample consisted of elementary school classroom teachers in grades one through five. The three outcome-based and one nonoutcome-based school districts were all in Iowa. Initial contacts were made by telephone with a follow-up letter to the district superintendent. The follow-up letter accomplished the following four tasks: 1) provided details of the study to the superintendent; 2) established written permission from the superintendent for the voluntary involvement of school personnel in the study; 3) verified school personnel numbers so questionnaire packets could be assembled and sent; and 4) established a point of contact in the district for future communication. Of five districts contacted, four agreed to participate in the study.

Once verification was received that a district was willing to participate, questionnaire packets were prepared and mailed to the previously identified point of contact. Each packet included the appropriate number of questionnaires and detailed directions for distribution, collection, and return of the completed questionnaires. In one district the initial group distribution method resulted in only three completed surveys out of a possible 26, so an individual, written appeal was made to the 23 teachers who did not return surveys, which resulted in 15 additional surveys being completed and returned.

The Iowa State University Committee on the Use of Human Subjects in Research reviewed this project and concluded that the rights and welfare
of the human subjects were adequately protected, that risks were outweighed by the potential benefits and expected value of the knowledge sought, that confidentiality of data was assured, and that consent was obtained by appropriate procedures.

Treatment of Data

The Chi-square test of homogeneity was identified as the appropriate statistical treatment of data following consultation with Professor Anton Netusil and other committee members. The chi-square test was used to compare the respondents from outcome- and nonoutcome-based schools because the use of ordinal data required a nonparametric test. The null hypotheses stated that responses would be independent of whether the teacher was in an outcome- or nonoutcome-based school at the .05 level of significance. Therefore, contingency tables were used to calculate the chi-square statistic to test assessment attitudes, practices, and concerns of teachers. Expected frequencies were calculated for each item so those with fewer than five could be identified. Once expected frequencies of fewer than five had been identified, a plan for joining cells was developed. Criteria considered when joining cells were the frequency count and similarity of responses.

In addition to the chi-square test, t-tests were used for the data in Section I, importance of assessment related classroom decisions. Statistical treatment of the data was accomplished using the software programs Statgraphics and Statview. The following chi-square formula was used for the test:
$\sum \frac{(O-E)^2}{E} = \chi^2$

$O =$ observed frequency $\quad E =$ expected frequencies

The first section of the questionnaire, teachers' attitudes concerning the importance of test related decisions, offered five scale options ranging from A=Important to E=Unimportant. Expected frequencies were calculated for the cells using the following formula. Several expected frequencies of fewer than five necessitated the joining of cells, which was successful for five of the eight items: response A=Important; response B=Somewhat important; and responses C, D, and E=Neutral/Unimportant.

Section II, practices of test use, offered four response options. Once again, expected frequencies of fewer than five necessitated the joining of cells. Responses A and B were joined to create a nonuse category; C=Effortful use; and D=Comfortable use. Joining of cells allowed analysis of three of the six items.

The third section of the questionnaire, teachers' concerns about different test types, originally offered seven response options. Again, in an effort to eliminate expected frequencies of fewer than five, cells were joined. Responses A, B, C, D, and E were joined to create a general concerns category; F=Using the tests more effectively; and G=No concerns. Joining cells allowed analysis of three of the six items.
While the joining of cells to eliminate expected frequencies of fewer than five was not successful with all items, further collapsing of cell categories would have led to a distortion of the data.

The degrees of freedom associated with this test was calculated using the following formula: \((R-1)(C-1)-(2-1)(3-1)=2\) degrees of freedom. The chi-square table shows a critical value of 5.991 for significance at the .05 level with 2 degrees of freedom. Thus, a chi-square statistic greater than 5.991 would cause the null hypotheses to be rejected. Because cells were joined to form three categories in each section, the degrees of freedom was 2 for each of the 20 individual item tests, so the same critical value of 5.991 was used to determine whether the null hypotheses were rejected.
CHAPTER IV. FINDINGS

Introduction

This study's major focus was whether implementation of the outcome-based school model was accompanied by a shift in teacher attitudes, practices, and concerns in the area of student assessment. Data were collected by using a 20-item questionnaire which was identified through a review of literature, adapted, and administered to grades one through five classroom teachers in four Iowa school districts.

In May and September of 1992, questionnaires for this study were sent to 153 teachers in three school districts that were classified as outcome-based, and 60 teachers in a district that was identified as nonoutcome-based. There was a potential of 213 total responses. A detailed analysis of each hypothesis appears immediately following the descriptive analysis of all returns.

Descriptive Analysis of All Returns

A total of 213 questionnaires were distributed to elementary school teachers in grades one through five in four Iowa school districts. Ninety-six completed questionnaires were returned, creating a return rate of 45 percent. Sixty-six of the 153 questionnaires distributed to teachers in outcome-based schools were completed and returned, representing a return rate of 43 percent. Thirty of the 60 questionnaires were returned from teachers in the nonoutcome-based school for a return
rate of 50 percent. Demographic information was collected in the following categories: gender, subject on which to focus, age of respondent, years of teaching experience, and education level. This information is displayed in Table 3.

Table 3. Demographics of respondents

<table>
<thead>
<tr>
<th>Gender</th>
<th>Outcome respondents</th>
<th>Nonoutcome respondents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>7 (10%)</td>
<td>5 (17%)</td>
<td>12 (13%)</td>
</tr>
<tr>
<td>Female</td>
<td>59 (89%)</td>
<td>25 (83%)</td>
<td>84 (88%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>Outcome respondents</th>
<th>Nonoutcome respondents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>36 (55%)</td>
<td>16 (53%)</td>
<td>52 (54%)</td>
</tr>
<tr>
<td>Mathematics</td>
<td>30 (45%)</td>
<td>14 (47%)</td>
<td>44 (46%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Outcome respondents</th>
<th>Nonoutcome respondents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 or under</td>
<td>4 (6%)</td>
<td>0 (0%)</td>
<td>4 (4%)</td>
</tr>
<tr>
<td>26-35 years</td>
<td>20 (30%)</td>
<td>5 (17%)</td>
<td>25 (26%)</td>
</tr>
<tr>
<td>36-45 years</td>
<td>23 (35%)</td>
<td>17 (57%)</td>
<td>40 (42%)</td>
</tr>
<tr>
<td>46 or older</td>
<td>19 (29%)</td>
<td>8 (27%)</td>
<td>27 (28%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Years of teaching</th>
<th>Outcome respondents</th>
<th>Nonoutcome respondents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1 year</td>
<td>4 (6%)</td>
<td>1 (3%)</td>
<td>5 (5%)</td>
</tr>
<tr>
<td>2-5 years</td>
<td>8 (12%)</td>
<td>6 (20%)</td>
<td>14 (15%)</td>
</tr>
<tr>
<td>6-12 years</td>
<td>16 (24%)</td>
<td>14 (47%)</td>
<td>30 (31%)</td>
</tr>
<tr>
<td>13-20 years</td>
<td>14 (21%)</td>
<td>9 (30%)</td>
<td>23 (24%)</td>
</tr>
<tr>
<td>more than 20 years</td>
<td>24 (36%)</td>
<td>0 (0%)</td>
<td>24 (25%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education level</th>
<th>Outcome respondents</th>
<th>Nonoutcome respondents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's degree</td>
<td>46 (70%)</td>
<td>15 (50%)</td>
<td>61 (64%)</td>
</tr>
<tr>
<td>Master's degree</td>
<td>16 (24%)</td>
<td>4 (13%)</td>
<td>20 (21%)</td>
</tr>
<tr>
<td>Master's + 30</td>
<td>4 (6%)</td>
<td>11 (37%)</td>
<td>15 (16%)</td>
</tr>
<tr>
<td>Doctorate</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>
Item Discrimination Questionnaire Analysis

A five-point scale was used for the items that probed teacher attitudes concerning the importance of eight test related decisions. The scale was presented on the questionnaire in this fashion:

IMPORTANT A B C D E UNIMPORTANT

Teachers were asked to use the scale to indicate the relative importance of each type of classroom decision for the subject area that had been specified (reading or math) and current grade level assignment.

A four-point scale was used for the items that probed teacher practices concerning use of six types of tests. Points one through four on the scale were presented on the questionnaire in this fashion:

A I do not currently use them and do not plan to use them in the future.
B I have decided to start using them in the future, but have not started to do so yet.
C I currently use them, but I find they take great effort to use.
D I use them on my own as a regular part of my instruction and do so comfortably.

Teachers were asked to indicate the statement that best described their current level of use of each type of test.

A seven-point scale was used for the items that probed teacher concerns about the use of each of the six types of tests. Points one
through seven on the scale were presented on the questionnaire in this fashion:

A  I am concerned about my lack of information about using this type of test.
B  I am concerned about my level of training, skill, and experience in using this type of test.
C  I am concerned about the amount of time required to manage the use of such tests.
D  I am concerned about how my students react when I administer this type of test.
E  I am concerned about establishing working relationships with other teachers using this type of test.
F  I am concerned about using these tests more effectively.
G  I have no concerns about using these tests more effectively.

Teachers were asked to indicate the one statement that best described their primary concern about the use of each of the six types of tests.

Importance of assessment related decisions

How important are different assessment related classroom decisions? The first eight questionnaire items were designed to answer that question. Subjects were asked to rate the importance of eight decisions on a five-point scale ranging from important to unimportant. Comparisons of the importance rating of decisions were used to determine whether those ratings were independent of whether the teachers were in an outcome-based
or nonoutcome-based school district. Contingency tables were used to calculate the chi-square statistic for testing independence at the .05 level of significance. For the purpose of analysis, response A=Important; response B=Somewhat important; and responses C, D, and E=Neutral/unimportant.

Treatment of these data revealed teachers' attitudes concerning the relative importance of each of the five types of educational assessment decisions where analysis was possible (items 2, 4, 6, 7, and 8), were independent of whether respondents were teaching in an outcome-based or nonoutcome-based school district. Expected cell counts of fewer than five made analysis of items 1, 3, and 5 impossible. Tables 4 through 11 display the results of the statistical analysis for the importance of the eight assessment related decisions.

Table 4. Importance of test information when diagnosing individual students' needs

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
<th>E</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important</td>
<td>O = 56</td>
<td>O = 24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E = 55</td>
<td>E = 25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somewhat important</td>
<td>O = 9</td>
<td>O = 5</td>
<td>E = 9.62</td>
<td>E = 4.38</td>
</tr>
<tr>
<td>Neutral/unimportant</td>
<td>O = 1</td>
<td>O = 1</td>
<td>E = 1.38</td>
<td>E = .62</td>
</tr>
</tbody>
</table>

*O=Observed frequencies; E=Expected frequencies.*
Though chi-square data are not provided for this item because three expected cell counts were fewer than five, a collective view of the results reveal that the majority of teachers (83%) from both types of schools classify test data as important for the purpose of diagnosing individual student needs.

Table 5. Importance of test information when diagnosing group needs

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important</td>
<td>O = 22</td>
<td>O = 11</td>
</tr>
<tr>
<td></td>
<td>E = 22.93</td>
<td>E = 10.07</td>
</tr>
<tr>
<td>Somewhat important</td>
<td>O = 32</td>
<td>O = 13</td>
</tr>
<tr>
<td></td>
<td>E = 31.26</td>
<td>E = 13.74</td>
</tr>
<tr>
<td>Neutral/unimportant</td>
<td>O = 12</td>
<td>O = 5</td>
</tr>
<tr>
<td></td>
<td>E = 11.81</td>
<td>E = 5.19</td>
</tr>
</tbody>
</table>

\[X^2 0.189 \quad p = 0.910\]

\(^a\)O=Observed frequencies; E=Expected frequencies.

While the chi-square statistic was not significant for the purpose of identifying differences between outcome and nonoutcome-based responses, data indicate that a large majority of teachers (82%) from both types of schools classify test information as important or somewhat important for the purpose of diagnosing group needs of students.
Table 6. Importance of test information when grouping students for instruction

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important</td>
<td>O = 11</td>
<td>O = 2</td>
</tr>
<tr>
<td></td>
<td>E = 9.03</td>
<td>E = 3.97</td>
</tr>
<tr>
<td>Somewhat important</td>
<td>O = 16</td>
<td>O = 8</td>
</tr>
<tr>
<td></td>
<td>E = 16.67</td>
<td>E = 7.33</td>
</tr>
<tr>
<td>Neutral/unimportant</td>
<td>O = 39</td>
<td>O = 19</td>
</tr>
<tr>
<td></td>
<td>E = 40.29</td>
<td>E = 17.71</td>
</tr>
</tbody>
</table>

*O=Observed frequencies; E=Expected frequencies.

Chi-square data for test information when grouping are not provided because one expected cell count was fewer than five. Collective analysis of the data indicate that a majority of teachers (61%) were neutral or considered test information as unimportant when grouping students for instruction.

Table 7. Importance of test information when determining achievement potential of students

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important</td>
<td>O = 13</td>
<td>O = 11</td>
</tr>
<tr>
<td></td>
<td>E = 16.67</td>
<td>E = 7.33</td>
</tr>
<tr>
<td>Somewhat important</td>
<td>O = 25</td>
<td>O = 9</td>
</tr>
<tr>
<td></td>
<td>E = 23.62</td>
<td>E = 10.38</td>
</tr>
<tr>
<td>Neutral/unimportant</td>
<td>O = 28</td>
<td>O = 9</td>
</tr>
<tr>
<td></td>
<td>E = 25.71</td>
<td>E = 11.29</td>
</tr>
<tr>
<td>X^2 3.586</td>
<td>p = 0.166</td>
<td></td>
</tr>
</tbody>
</table>

*O=Observed frequencies; E=Expected frequencies.
Results of this analysis are puzzling. Test information should be very important for determining outcome-based education achievement potential. Though the chi-square statistic was not significant, combined data show almost an equal distribution of responses between the three response categories.

Table 8. Importance of test results when assigning grades

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important</td>
<td>$O^a = 6$</td>
<td>$O = 4$</td>
</tr>
<tr>
<td></td>
<td>$E^a = 6.84$</td>
<td>$E = 3.16$</td>
</tr>
<tr>
<td>Somewhat important</td>
<td>$O = 15$</td>
<td>$O = 8$</td>
</tr>
<tr>
<td></td>
<td>$E = 15.74$</td>
<td>$E = 7.26$</td>
</tr>
<tr>
<td>Neutral/unimportant</td>
<td>$O = 44$</td>
<td>$O = 18$</td>
</tr>
<tr>
<td></td>
<td>$E = 42.42$</td>
<td>$E = 19.58$</td>
</tr>
</tbody>
</table>

$^aO=$Observed frequencies; $E=$Expected frequencies.

Chi-square data are not provided in Table 8 because one expected cell count was fewer than five. Combined results for respondents from both types of schools reveal that more than 60 percent of teachers in this study had a neutral attitude or considered test information as unimportant for the purpose of assigning grades. Less than 10 percent responded that test results were important for grading purposes.
Table 9. Importance of test information when evaluating an instructional unit

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important</td>
<td>O = 29</td>
<td>0 = 15</td>
</tr>
<tr>
<td></td>
<td>E = 30.11</td>
<td>E = 13.89</td>
</tr>
<tr>
<td>Somewhat important</td>
<td>O = 26</td>
<td>0 = 9</td>
</tr>
<tr>
<td></td>
<td>E = 23.95</td>
<td>E = 11.05</td>
</tr>
<tr>
<td>Neutral/unimportant</td>
<td>O = 10</td>
<td>0 = 6</td>
</tr>
<tr>
<td></td>
<td>E = 10.95</td>
<td>E = 5.05</td>
</tr>
</tbody>
</table>

\[ X^2 = 0.945 \quad p = 0.623 \]

*O=Observed frequencies; E=Expected frequencies.

The chi-square statistic in Table 9 was not significant. Combined results show that approximately half of responding teachers (46%) consider test results as important when determining the success of an instructional unit.

Table 10. Importance of test information to communicate academic expectations

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important</td>
<td>O = 28</td>
<td>0 = 14</td>
</tr>
<tr>
<td></td>
<td>E = 28.74</td>
<td>E = 13.26</td>
</tr>
<tr>
<td>Somewhat important</td>
<td>O = 25</td>
<td>0 = 9</td>
</tr>
<tr>
<td></td>
<td>E = 23.26</td>
<td>E = 10.74</td>
</tr>
<tr>
<td>Neutral/unimportant</td>
<td>O = 12</td>
<td>0 = 7</td>
</tr>
<tr>
<td></td>
<td>E = 13.00</td>
<td>E = 6.00</td>
</tr>
</tbody>
</table>

\[ X^2 = 0.714 \quad p = 0.700 \]

*O=Observed frequencies; E=Expected frequencies.
49

Though the chi-square statistic was not significant in Table 10, combined results show that almost half of responding teachers (44%) consider tests to be important tools to use when communicating expectations to students.

Table 11. Importance of tests to control and motivate students

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important</td>
<td>O = 30</td>
<td>O = 15</td>
</tr>
<tr>
<td></td>
<td>E = 31.26</td>
<td>E = 13.74</td>
</tr>
<tr>
<td>Somewhat important</td>
<td>O = 18</td>
<td>O = 9</td>
</tr>
<tr>
<td></td>
<td>E = 18.76</td>
<td>E = 8.24</td>
</tr>
<tr>
<td>Neutral/unimportant</td>
<td>O = 18</td>
<td>O = 5</td>
</tr>
<tr>
<td></td>
<td>E = 15.98</td>
<td>E = 7.02</td>
</tr>
</tbody>
</table>

\[ X^2 1.105 \quad p = 0.576 \]

*O=Observed frequencies; E=Expected frequencies.

The chi-square statistic in Table 11 was not significant. Combined responses show that almost half of teachers (47%) consider tests an important tool to control and motivate students.

Level of use of different assessment types

What is the level of use of different types of assessment? The six items in Section II of the questionnaire were designed to answer that question. Subjects were asked to identify the level of use of six different types of assessment. Levels of use were nonuse, planned use,
effortful use, and regular use. Comparison of the data was used to
determine whether elementary school teachers' testing practices were
independent of whether they were in an outcome-based or nonoutcome-based
school district. Contingency tables were used to calculate the chi-square
statistic for testing independence at the .05 level of significance. For
the purpose of analysis, responses A and B=Nonuse; response C=Use with
effort; and response D=Comfortable use.

Analysis of these data revealed teachers' test use of the three types
of educational tests where analysis was possible (items 9, 11, and 13),
were independent of whether respondents were teaching in an outcome- or
nonoutcome-based school district. Results for items 10, 12, and 13 could
not be analyzed due to expected cell counts of fewer than five. Tables 12
through 17 display the level of use results for each of the six assessment
types.

Table 12. Teachers' level of use of objective teacher-made paper and
pencil tests

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonuse</td>
<td>O^a = 20</td>
<td>O = 8</td>
</tr>
<tr>
<td></td>
<td>E^a = 19.57</td>
<td>E = 8.43</td>
</tr>
<tr>
<td>Effortful use</td>
<td>O = 12</td>
<td>O = 7</td>
</tr>
<tr>
<td></td>
<td>E = 13.28</td>
<td>E = 5.72</td>
</tr>
<tr>
<td>Comfortable use</td>
<td>O = 33</td>
<td>O = 13</td>
</tr>
<tr>
<td></td>
<td>E = 32.15</td>
<td>E = 13.85</td>
</tr>
</tbody>
</table>

\[ X^2 \] 0.515 \quad p = 0.773

^aO=Observed frequencies; E=Expected frequencies.
While the chi-square statistic was not significant for Table 12, collective analysis of the data reveal that almost half of responding teachers (49%) report that they comfortably use objective tests that they design themselves. Almost a third of respondents (30%) reported that they did not use personally developed tests.

Table 13. Teachers' level of use of text embedded tests

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$O^a$ = 15</td>
<td>$O = 0$</td>
</tr>
<tr>
<td></td>
<td>$E^a$ = 10.37</td>
<td>$E = 4.63$</td>
</tr>
<tr>
<td>Effortful use</td>
<td>$O = 15$</td>
<td>$O = 10$</td>
</tr>
<tr>
<td></td>
<td>$E = 17.29$</td>
<td>$E = 7.71$</td>
</tr>
<tr>
<td>Comfortable use</td>
<td>$O = 35$</td>
<td>$O = 19$</td>
</tr>
<tr>
<td></td>
<td>$E = 37.34$</td>
<td>$E = 16.67$</td>
</tr>
</tbody>
</table>

$^a$O=Observed frequencies; E=Expected frequencies.

Chi-square data are not provided for Table 13 because one expected cell count was fewer than five. More than half of all teachers responded (57%) that they comfortably use tests that are included with textbooks and other instructional materials.
Table 14. Teachers' level of use of standardized achievement tests

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonuse</td>
<td>O = 21</td>
<td>O = 9</td>
</tr>
<tr>
<td></td>
<td>E = 20.65</td>
<td>E = 9.35</td>
</tr>
<tr>
<td>Effortful use</td>
<td>O = 25</td>
<td>O = 11</td>
</tr>
<tr>
<td></td>
<td>E = 24.77</td>
<td>E = 11.23</td>
</tr>
<tr>
<td>Comfortable use</td>
<td>O = 18</td>
<td>O = 9</td>
</tr>
<tr>
<td></td>
<td>E = 18.58</td>
<td>E = 8.42</td>
</tr>
</tbody>
</table>

X² 0.084  \[ p = 0.959 \]

*O=Observed frequencies; E=Expected frequencies.

Though the chi-square statistic in Table 14 was not significant, collective results show that teachers are almost equally divided in their level of use of standardized achievement tests. Approximately a third each reported that they comfortably use (32%), use with special effort (39%), or do not use (29%) standardized achievement tests.

Table 15. Teachers' level of use of oral questioning

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonuse</td>
<td>O = 2</td>
<td>O = 0</td>
</tr>
<tr>
<td></td>
<td>E = 1.38</td>
<td>E = .62</td>
</tr>
<tr>
<td>Effortful use</td>
<td>O = 2</td>
<td>O = 2</td>
</tr>
<tr>
<td></td>
<td>E = 2.77</td>
<td>E = 1.23</td>
</tr>
<tr>
<td>Comfortable use</td>
<td>O = 61</td>
<td>O = 27</td>
</tr>
<tr>
<td></td>
<td>E = 60.85</td>
<td>E = 27.15</td>
</tr>
</tbody>
</table>

*O=Observed frequencies; E=Expected frequencies.
Chi-square data are not provided for Table 15 because one expected cell count was fewer than five. Combined results for all teachers indicate that almost all teachers (94%) are comfortable using oral questioning of students as a means of assessing student learning.

Table 16. Teachers' level of use of structured performance assessment

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonuse</td>
<td>$O^8 = 14$</td>
<td>$O = 6$</td>
</tr>
<tr>
<td></td>
<td>$E^8 = 13.70$</td>
<td>$E = 6.30$</td>
</tr>
<tr>
<td>Effortful use</td>
<td>$O = 17$</td>
<td>$O = 11$</td>
</tr>
<tr>
<td></td>
<td>$E = 19.17$</td>
<td>$E = 8.83$</td>
</tr>
<tr>
<td>Comfortable use</td>
<td>$O = 32$</td>
<td>$O = 12$</td>
</tr>
<tr>
<td></td>
<td>$E = 30.13$</td>
<td>$E = 13.87$</td>
</tr>
<tr>
<td>$X^2 = 1.171$</td>
<td>$p = 0.557$</td>
<td></td>
</tr>
</tbody>
</table>

$^8O$=Observed frequencies; $E$=Expected frequencies.

While the chi-square statistic was not significant for structured performance assessment, combined results show that almost half of the teachers (48%) are comfortable using structured performance assessments. A structured performance was described in the questionnaire as those performances where 1) a clearly defined reason for the assessment has been given, 2) an activity is planned to elicit responses from the student, 3) an evaluation plan for responses has been specified, and 4) scoring procedures are communicated to students.
Table 17. Teachers' level of use of spontaneous performance assessment

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O = 5</td>
<td>O = 1</td>
</tr>
<tr>
<td></td>
<td>E = 4.15</td>
<td>E = 1.85</td>
</tr>
<tr>
<td>Nonuse</td>
<td>O = 5</td>
<td>O = 5</td>
</tr>
<tr>
<td></td>
<td>E = 6.91</td>
<td>E = 3.09</td>
</tr>
<tr>
<td>Effortful use</td>
<td>O = 55</td>
<td>O = 23</td>
</tr>
<tr>
<td></td>
<td>E = 53.94</td>
<td>E = 24.06</td>
</tr>
</tbody>
</table>

*O-Observed frequencies; E-Expected frequencies.

Chi-square data are not provided for Table 17 because three expected cell counts were fewer than five. Most teachers (84%) reported that they comfortably use spontaneous performances as a means of evaluating student achievement. Spontaneous performances are defined as those activities that offer informal opportunities for the teacher to observe and evaluate a student's performance and to judge the student's proficiency.

Assessment concerns

What concerns are related to different types of assessments? The six items in Section III of the questionnaire were developed to answer that question. Teachers identified a stage of concern for each of six types of assessment. Comparisons were made for the three items where analysis was possible, to determine whether elementary school teachers' concerns about tests were independent of whether they were in an outcome-based or nonoutcome-based school district. Contingency tables were used to
calculate the chi-square statistic for testing independence at the .05 level of significance. For the purpose of analysis, responses A, B, C, D, and E = General concerns; response F = Effectiveness concerns; and response G = No concerns.

Analysis of these data revealed teachers' concerns about each of the three different types of educational tests where analysis was possible (items 16, 19, and 20), were independent of whether they were teaching in an outcome- or nonoutcome-based school district. Analysis of items 15, 17, and 18 was not possible due to an expected cell size of fewer than five. Tables 18 through 23 display the assessment concern results for each of the six types of assessment.

Table 18. Teachers' concerns about their use of objective teacher-made paper and pencil tests

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>General concerns</td>
<td>O = 18</td>
<td>O = 10</td>
</tr>
<tr>
<td></td>
<td>E = 19.60</td>
<td>E = 8.40</td>
</tr>
<tr>
<td>Effectiveness concerns</td>
<td>O = 11</td>
<td>O = 5</td>
</tr>
<tr>
<td></td>
<td>E = 11.20</td>
<td>E = 4.80</td>
</tr>
<tr>
<td>No concerns</td>
<td>O = 34</td>
<td>O = 12</td>
</tr>
<tr>
<td></td>
<td>E = 32.2</td>
<td>E = 13.80</td>
</tr>
</tbody>
</table>

*O=Observed frequencies; E=Expected frequencies.*

Chi-square data are not provided for Table 18 because one expected cell count was fewer than five. Combined results of all responses reveal that more than half of the teachers (51%) expressed no worries about using objective paper and pencil tests that they had developed. Almost a third
of the responding teachers (31%) expressed general concerns such as a lack of information or training about developing these types of tests, student reactions to the tests, and finding ways to collaborate with other teachers to design and use locally developed tests. Nearly a fifth of respondents (18%) indicated that they were most interested in refining their own tests to make them more effective.

Table 19. Teachers' concerns about their use of text embedded tests

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>General concerns</td>
<td>O = 22</td>
<td>O = 6</td>
</tr>
<tr>
<td></td>
<td>E = 19.21</td>
<td>E = 8.79</td>
</tr>
<tr>
<td>Effectiveness concerns</td>
<td>O = 15</td>
<td>O = 10</td>
</tr>
<tr>
<td></td>
<td>E = 17.15</td>
<td>E = 7.85</td>
</tr>
<tr>
<td>No concerns</td>
<td>O = 22</td>
<td>O = 11</td>
</tr>
<tr>
<td></td>
<td>E = 23.64</td>
<td>E = 10.36</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 2.208 \quad p = 0.331 \]

*O=Observed frequencies; E=Expected frequencies.

The chi-square statistic was not significant in Table 19. Summary figures show that responses were approximately the same between the three concerns categories of general (33%), effectiveness (29%), and those having no concerns (38%) about the use of tests included in textbooks and other instructional materials.
Table 20. Teachers' concerns about their use of standardized achievement tests

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>General concerns</td>
<td>$O^a = 43$</td>
<td>$O = 17$</td>
</tr>
<tr>
<td></td>
<td>$E^a = 40.69$</td>
<td>$E = 19.31$</td>
</tr>
<tr>
<td>Effectiveness concerns</td>
<td>$O = 5$</td>
<td>$O = 5$</td>
</tr>
<tr>
<td></td>
<td>$E = 6.78$</td>
<td>$E = 3.22$</td>
</tr>
<tr>
<td>No concerns</td>
<td>$O = 11$</td>
<td>$O = 6$</td>
</tr>
<tr>
<td></td>
<td>$E = 11.53$</td>
<td>$E = 5.47$</td>
</tr>
</tbody>
</table>

$^a$O=Observed frequencies; E=Expected frequencies.

Though chi-square data are not provided for Table 20 because four expected cell counts were fewer than five, collective analyses of responses show that more than 80 percent of teachers expressed some type of concern about their use of standardized achievement tests. More than two-thirds of respondents (69%) expressed general concerns such as

Table 21. Teachers' concerns about their use of oral questioning

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>General concerns</td>
<td>$O^a = 1$</td>
<td>$O = 0$</td>
</tr>
<tr>
<td></td>
<td>$E^a = 0.70$</td>
<td>$E = 0.30$</td>
</tr>
<tr>
<td>Effectiveness concerns</td>
<td>$O = 7$</td>
<td>$O = 3$</td>
</tr>
<tr>
<td></td>
<td>$E = 7.03$</td>
<td>$E = 2.97$</td>
</tr>
<tr>
<td>No concerns</td>
<td>$O = 56$</td>
<td>$O = 24$</td>
</tr>
<tr>
<td></td>
<td>$E = 56.26$</td>
<td>$E = 23.74$</td>
</tr>
</tbody>
</table>

$^a$O=Observed frequencies; E=Expected frequencies.
training, time expenditure, and student reactions to these tests, with only 11 percent expressing anxiety about refining their use of standardized tests to make them more effective.

While chi-square data are not provided for use of oral questioning because four expected cell counts were fewer than five, summary information suggests that few teachers (12%) have any concerns about their ability to effectively use oral questioning as a means of assessing student achievement.

Table 22. Teachers' concerns about their use of structured performance assessment

<table>
<thead>
<tr>
<th>Responses</th>
<th>Outcome-based</th>
<th>Nonoutcome-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>General concerns</td>
<td>O = 31</td>
<td>O = 10</td>
</tr>
<tr>
<td></td>
<td>E = 28.84</td>
<td>E = 12.16</td>
</tr>
<tr>
<td>Effectiveness concerns</td>
<td>O = 14</td>
<td>O = 4</td>
</tr>
<tr>
<td></td>
<td>E = 12.66</td>
<td>E = 5.34</td>
</tr>
<tr>
<td>No concerns</td>
<td>O = 19</td>
<td>O = 13</td>
</tr>
<tr>
<td></td>
<td>E = 22.51</td>
<td>E = 9.49</td>
</tr>
</tbody>
</table>

\[ X^2 \ 2.867 \quad p = 0.239 \]

*O=Observed frequencies; E=Expected frequencies.

The chi-square statistic was not significant for identifying differences between teachers in outcome- and nonoutcome-based schools. Approximately a third of responding teachers (35%) stated that they had no concerns about their ability to develop and effectively use structured performances as a way to assess student performance, with the remaining
two-thirds expressing some general concerns or anxiety about using performance evaluations more effectively.

Though the chi-square statistic was not significant for Table 23, summary data indicate that almost two-thirds of the teachers (62%) expressed no concerns about their ability to effectively use student performance information gathered informally during classroom activities.

Because no significant differences were found, additional ways to treat and analyze the data were designed in order to learn as much as possible from the data. First, outcome-based and nonoutcome-based responses were grouped together. Then the subject on which respondents were to focus was used to disaggregate the data. Frequency counts, contingency tables, expected cell counts, and chi-square were created for
each item. Responses for reading and those for mathematics were compared. No significant differences were identified except for item seven. Table 24 displays the results of the importance of communicating academic expectations.

Table 24. Importance of tests for communicating expectations in reading and mathematics

<table>
<thead>
<tr>
<th>Responses</th>
<th>Reading</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important</td>
<td>O = 24</td>
<td>O = 18</td>
</tr>
<tr>
<td></td>
<td>E = 22.5474</td>
<td>E = 19.4526</td>
</tr>
<tr>
<td>Somewhat important</td>
<td>O = 23</td>
<td>O = 11</td>
</tr>
<tr>
<td></td>
<td>E = 18.2526</td>
<td>E = 15.7474</td>
</tr>
<tr>
<td>Neutral/unimportant</td>
<td>O = 4</td>
<td>O = 15</td>
</tr>
<tr>
<td></td>
<td>E = 10.2000</td>
<td>E = 8.8000</td>
</tr>
</tbody>
</table>

\[ X^2 = 6.71880, \quad p = .0347562^* \]

^O=Observed frequencies; E=Expected frequencies.

^*Significant at the .05 level.

A statistically significant difference was found concerning the use of tests to communicate expectations when teachers were asked to focus on reading or mathematics. Teachers who were asked to focus on reading reported that testing as a means of communicating expectations was less important than did teachers of mathematics.

Data were disaggregated using the demographic categories of age and education level of respondents. Significant differences were found for one item when data were disaggregated by age. Table 25 shows the results of that item.
Table 25. Use of standardized achievement tests by age categories

<table>
<thead>
<tr>
<th>Responses</th>
<th>35 and younger</th>
<th>36 and older</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonuse</td>
<td>O(^a) = 13</td>
<td>O = 17</td>
</tr>
<tr>
<td></td>
<td>E(^a) = 9.3548</td>
<td>E = 20.9677</td>
</tr>
<tr>
<td>Effortful use</td>
<td>O = 11</td>
<td>O = 25</td>
</tr>
<tr>
<td></td>
<td>E = 10.8387</td>
<td>E = 25.1613</td>
</tr>
<tr>
<td>Comfortable use</td>
<td>O = 4</td>
<td>O = 23</td>
</tr>
<tr>
<td></td>
<td>E = 8.1290</td>
<td>E = 18.8710</td>
</tr>
</tbody>
</table>

X\(^2\) 7.49686  p = .0235547

\(^a\)O=Observed frequencies; E=Expected frequencies.

More teachers aged 36 and older reported that they were comfortable using standardized tests in the classroom than did respondents who were younger.

A significant difference was also found for item 13 when the data were disaggregated using the education level of the respondent. Table 26 shows the results of that analysis.

A statistically significant difference was found with regard to the use of structured performance assessment based on the education level of the respondent. Teachers with a bachelor's degree more often reported that they comfortably used structured performance assessment than did those respondents with a master's degree or higher.
Table 26. Use of structured performance assessment by education level

<table>
<thead>
<tr>
<th>Responses</th>
<th>Bachelor's</th>
<th>Master’s or higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonuse</td>
<td>O = 16</td>
<td>O = 4</td>
</tr>
<tr>
<td></td>
<td>E = 13.0435</td>
<td>E = 6.9565</td>
</tr>
<tr>
<td>Effortful use</td>
<td>O = 12</td>
<td>O = 16</td>
</tr>
<tr>
<td></td>
<td>E = 18.2609</td>
<td>E = 9.7391</td>
</tr>
<tr>
<td>Comfortable use</td>
<td>O = 32</td>
<td>O = 12</td>
</tr>
<tr>
<td></td>
<td>E = 28.6957</td>
<td>E = 14.8211</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 10.4308 \quad p = .00543215^{**} \]

\(^{a}\text{O-Observed frequencies; E-Expected frequencies.}\)

\(^{**}\text{Significant at the .01 level.}\)
CHAPTER V. SUMMARY, CONCLUSIONS, LIMITATIONS, DISCUSSION, AND RECOMMENDATIONS

Summary

This study probed the assessment attitudes, practices, and concerns of classroom teachers in outcome and nonoutcome-based schools. A total of 96 Iowa elementary school teachers in grades one through five participated in this study about assessment. Sixty-six teachers in outcome-based schools and 30 teachers in nonoutcome-based schools completed the 20-item questionnaire. Section I consisted of eight items that probed teachers' attitudes about assessment related decisions using a five-point scale. Section II consisted of six items to determine the level of use for different types of assessment instruments. Section III items were used to determine the primary concerns of teachers about six types of assessment instruments.

The chi-square test was used to analyze the data to determine whether teachers' attitudes, practices, and concerns were independent of whether they were teaching in an outcome- or nonoutcome-based school district. Frequency counts and contingency tables were used to determine the chi-square statistic and probability for testing responses to each item at the .05 level of significance.

A related purpose of this study was to examine classroom assessment practices using demographic information. Post hoc analyses were conducted to determine whether responses were independent of the subject on which the respondent was asked to focus, age, and education level of the respondent.
Hypothesis 1

Teachers' attitudes concerning the relative importance of different types of educational assessment decisions are independent of whether the elementary school teachers are in an outcome- or nonoutcome-based school district in Iowa.

Expected frequency counts of five or more for each cell were needed to proceed with the chi-square test. Analysis of the data was possible for five of the eight items (2, 4, 6, 7, and 8). The null hypothesis was accepted for all five of the items when data were disaggregated based on membership in an outcome- or nonoutcome-based school district. Teachers in both types of districts rated the importance of the following assessment related decisions the same: diagnosing group needs; determining the achievement potential of students; evaluating the instructional unit to see if it worked; communicating academic expectations; and controlling and motivating students.

When data for all teachers were disaggregated based on the subject on which the respondent was asked to focus, age, and education level, one significant difference was found. Teachers who were asked to focus on reading reported that testing as a means of communicating expectations was less important than did teachers of mathematics.

Combined results for the relative importance of each type of assessment related classroom decision reveal that diagnosing individual needs of students is the most important purpose of classroom assessment. Teachers also reported that diagnosing group needs, evaluating an instructional unit, communicating academic expectations, and motivating
students were also important decisions related to assessment. Grouping students for instruction, determining achievement potential, and assigning grades were viewed as less important purposes of assessment.

Hypothesis 2

Assessment practices are independent of whether the elementary school teacher is in an outcome- or nonoutcome-based school district in Iowa.

Analysis of the data was possible for three of the six items (9, 11, and 13) because expected cell frequencies for all cells were five or more. The null hypotheses were retained for all three items when data were disaggregated based on teachers' membership in an outcome- or nonoutcome-based district. Teachers in both types of schools rated their use of the following types of tests the same: objective teacher-made paper and pencil tests; standardized achievement tests; and structured performance assessment. Analysis of data for text embedded tests, oral questioning, and spontaneous performance assessment was not possible because one or more expected cell frequency was less than five.

When data for all teachers were disaggregated based on the subject on which the respondent was asked to focus, age, and education level, two statistically significant differences were found. More teachers aged 36 and older reported that they were comfortable using standardized tests in the classroom than did respondents who were younger. A difference was also found with regard to the use of structured performance assessment. When the data were disaggregated based on education level, more teachers
with a bachelor's degree reported that they comfortably used structured performance assessment than did those with a master's degree or higher.

Combined data for all teachers concerning their use of different assessment types reveal that they are most comfortable using oral questioning and spontaneous performance assessment to monitor student achievement. About half of the teachers reported that they comfortably use text embedded tests, teacher-made tests, and structured performance assessment. Standardized achievement tests were the type of assessment that the fewest number of teachers reported they were comfortable using.

**Hypothesis 3**

Assessment concerns are independent of whether the elementary school teacher is in an outcome- or nonoutcome-based school district in Iowa.

Analysis of the data was possible for three of the six items (16, 19, and 20) relating to teachers' concerns about different assessment types, with the null hypothesis being accepted for each. Teachers' concerns about their ability to use text embedded tests, structured performance assessment, and spontaneous performance assessment were independent of their membership in an outcome- or nonoutcome-based school district. Analysis of the data for teacher-made tests, standardized achievement tests, and oral questioning was not possible due to one or more expected cell frequencies of fewer than five.

When data for all teachers were disaggregated based on the subject on which the respondent was asked to focus, age, and education level, no significant differences were identified.
Combined results show that most teachers have no concerns about their ability to use oral questioning, spontaneous performance assessment, and teacher-made tests to measure student achievement. Approximately two-thirds of the teachers expressed some type of concern about text embedded tests and structured performance assessment. Causing the greatest concern was the use of standardized achievement tests. Four out of five teachers expressed some kind of concern about standardized tests.

Analysis of Data

1. Eleven of the 20 items were tested using the chi-square statistic with no significant differences found between the respondents in outcome- and nonoutcome-based schools.

2. Analysis of outcome- and nonoutcome-based responses for nine of the responses could not be accomplished as a result of expected cell counts of fewer than five.

3. One statistically significant difference was found when data were disaggregated by the subject on which teachers were asked to focus when responding to the 20 items. Reading teachers reported that testing as a means of communicating expectations was less important than did teachers of mathematics.

4. One statistically significant difference was identified when data were disaggregated by age. More teachers aged 36 and older reported that they are comfortable using standardized tests in the classroom.
5. One statistically significant difference was identified when data were disaggregated based on the educational level of the respondent. Teachers with a bachelor's degree more often reported that they comfortably use structured performance assessment than teachers with a master's degree or higher education level.

Conclusions

The following conclusions are offered concerning the analysis of the data and compilation of information collected in the review of literature.

1. It appears that no real differences exist between assessment attitudes, practices, and concerns for teachers in outcome- and nonoutcome-based schools.

2. "Diagnosing strengths and weaknesses of individual students" is viewed as the most important purpose of assessment for teachers in both outcome- and nonoutcome-based schools.

3. "Assigning grades to students" is viewed as the least important purpose of assessment.

4. Teachers are most comfortable using oral questioning and spontaneous performances as means for assessing student achievement.

5. Most teachers have no concerns about their ability to use oral questioning and spontaneous performances to assess student achievement.
Limitations

While efforts were made to ensure that this study was rigorous and made a worthwhile contribution to educational research, the following limitations must be noted:

1. No attempt was made to determine the effectiveness of the implementation in the three outcomes districts.
2. No attempt was made to determine whether assessment attitudes about assessment related decisions, assessment use, and concerns about assessment types had been a focus during outcomes implementation.
3. Each school district participating in this investigation did so on a voluntary basis. The decision to take part may indicate a special interest on the part of the district that could have influenced the results.
4. Participation in this study was voluntary on the part of teachers. This decision may have influenced the results.
5. The investigation focused on elementary school teachers in grades one through five. Results may not be generalizable to other grade levels.
6. This study asked teachers to respond to questions in terms of reading or mathematics. Results may not be consistent with other subject areas.
7. The school districts that participated in the study were all from Iowa. Results may be inconsistent with those conducted in other parts of the nation.
8. The investigation relied on self-reporting of attitudes, practices, and concerns of teachers. No attempt was made to confirm whether responses were consistent with genuine teacher actions.

9. Outcome-based school restructuring was the only educational reform movement explored. Results may not generalize to other restructuring efforts.

10. Small numbers of subjects may have affected the significance of differences between the groups. Larger numbers may have changed some of the results.

11. The outcomes districts were in the early stages of implementing outcomes education. Results may vary as the districts move more completely into outcome-based education.

Discussion

The purpose of this study was to determine whether the transformational changes that OBE was designed to produce were realized. Was OBE really working in the schools where it had been selected as the model for restructuring? This study sought to use teachers' assessment attitudes, practices, and concerns as a measure of change to provide insight as to whether implementation of OBE resulted in fundamental change in the structure of schooling.

The results indicate that no real differences exist between student assessment in the OBE schools and the nonoutcome school that participated in this study, suggesting that the basic school structure has remained the
same. Five key conclusions can be identified that may explain the lack of differences found.

The most optimistic conclusion is that the nonoutcome-based school, while not officially part of the OBE movement, is already philosophically aligned with the OBE model. The nonoutcome school in this study may have quietly accomplished what the official, more conspicuous OBE schools are working toward. They may be designing, developing, delivering, and documenting the educational process in relation to the identified outcomes. They may have shifted away from time and curriculum based schooling to one where clearly defined student outcomes form the basis for all instructional decisions. When OBE is fully implemented, it completely changes the way schools function. Members of the nonoutcome district may have recognized the key role tests play in instruction and are effectively using them to set the standards for the school.

A second possible conclusion is that the OBE districts in this study have not fully implemented the model or have implemented it in a superficial way. It is much easier to attach a label, such as OBE, to a district, than it is to actually implement the necessary changes. The shift from traditional, content, and time based instruction to student outcome-based schooling requires a change in how learning is assessed. Paper and pencil, multiple-choice tests, long the backbone classroom assessment, are inappropriate when documenting the more complex types of learning associated with OBE. Carefully structured student performance tasks that replicate effective adult behavior will need to become a significant part of any OBE assessment system. Teachers need to be taught
how to design and use alternative types of student assessment instruments. They must receive ongoing support to successfully accomplish the shift in student assessment.

The third major conclusion is that more time is required before the significant differences between OBE and nonoutcome assessment practices can be accurately measured. Each of the OBE schools had been associated with restructuring approximately three years. The magnitude of the shift to OBE may be such that five to ten years will be required for full implementation. The major student assessment changes required by OBE may only be evident several years in the future. Real change takes time.

A fourth conclusion is that the 20-item questionnaire used is not a valid measure of differences in teachers' classroom assessment practices. The questionnaire used was developed from a much longer instrument and the elimination of certain items may have invalidated the instrument.

The fifth and final conclusion is that OBE, as practiced, does not make a difference in teachers' assessment practices. The transformational nature of OBE may be an illusion rather than a reality.

Any approach to changing schools will begin with a set of assumptions about schools (Wiles, 1993). One fundamental assumption schools moving into OBE will need to accept is that student tests not only monitor standards, but also set them (Wiggins, 1989). Adoption of that assumption early in the implementation stage allows educators to address the issue of student assessment from the beginning, rather than as an afterthought or not at all. Attempts to change instruction and educational content will be negated when progress is measured using tests of the past.
If genuine change in American education is the goal, reform efforts need to look at student assessment practices as a starting point. Redesigning how we measure student success is the essential first step in the recreation of the public system of education.

Recommendations for Practice

The following recommendations for practice are offered in particular to school districts that are planning to implement OBE.

1. Prior to implementing OBE, it is important to thoroughly analyze the current state of all aspects of the school district’s programs. It will be important to include assessment attitudes, practices, and concerns as key components to be analyzed. The resulting data can be used to reflect on conditions from the past that have led to the current state as well as serve to initiate discussion about the direction the district intends to move. The data can also be used to firmly establish a need for change as well as create a bridge from the past to the desired future. The movement toward OBE can then be viewed in context and become a positive extension of the past.

2. After the need for change is established, it is necessary to create a clear vision of what an OBE district and the accompanying OBE assessment practices look like. Once the vision is clarified, plans for the necessary staff development can be made. Teachers will need to have many opportunities to learn about and experiment with alternatives to the traditional types
of assessment instruments. Teachers will need to learn about alternative assessment tools such as scoring rubrics, portfolios, exhibitions, and authentic performance tasks. One-time assessment training will not be enough to bring about the desired changes; therefore, a strategic plan for ongoing staff development will be needed.

3. When OBE assessment practices have been identified and teachers have had the necessary training, implementation will begin. Real change takes place over time—often three or more years will be needed. Planning for change by stages will establish an organized structure for that change to take place and will create manageable, short-term goals that will lead to total implementation of OBE. Well-planned, short-term goals also allow early successes to be realized. Throughout the implementation stages it is important to recognize the need for continuing support and training.

4. As a district implements OBE assessment practices, it is important to monitor and evaluate the process and progress toward the identified goals. It is a way of validating that the change to OBE was worth the effort. A well-planned evaluation will document for teachers as well as for community members that OBE was the right choice. Involving teachers and community members in the evaluation design and data analysis will help to establish a partnership between the school and the community.
Recommendations for Further Research

Based on the findings and conclusions of this study, the following recommendations for further research are offered.

1. A related study should be conducted to ascertain the extent to which OBE was actually implemented by classroom teachers. The Level of Use of an innovation (Hall, 1984) or similar diagnostic tool should be used to document the teacher behaviors that can be used to establish the extent to which OBE implementation has taken place in the classroom.

2. A related study should be conducted to determine whether the non-significant differences found were due to a lack of focus on classroom assessment during the implementation of OBE. Was classroom assessment included as a significant component when the teachers developed or received training in OBE?

3. A similar research study should be conducted that includes a larger number of OBE districts. An established network such as the Association for Supervision and Curriculum Development, National Educational Goals Panel, National Governor's Association, or the National Assessment Governing Board can serve to increase the sample size and create a more reliable picture of OBE schools and districts.

4. A research study should be initiated that focuses on assessment attitudes, practices, and concerns of middle and high school teachers in OBE districts and schools. Results from elementary teachers may differ from those of teachers at other grade levels.
5. A study should be conducted that probes OBE classroom assessment practices for all content areas. If assessment practices in specific content areas are aligned with appropriate OBE practices, those areas can serve as the foundation for aligning assessment in other areas.

6. A more comprehensive study should be conducted that consists of a cross-section of schools across the United States. Iowa schools offer too narrow a sample for accurate generalizations to be made.

7. A qualitative study should be designed to investigate teacher assessment behaviors in the classroom. Observational data can provide a more comprehensive, accurate profile of actual behaviors than teachers' self-reporting of attitudes, practices, and concerns.

8. A comprehensive study of other school reform efforts such as site-based management, cooperative learning, and integrating the curriculum should be initiated to research the success of each. The data may lead to identification of the essential components for successful educational reform implementation.

9. A longitudinal research study should be conducted to determine whether the desired assessment attitudes, practices, and concerns are developed in later years of OBE implementation. Teachers' classroom assessment behaviors should be systematically monitored throughout the implementation time span. If a predictable
pattern of change can be identified, it may lead to a smoother transition during the period of implementation.

Epilogue

In order to ascertain the current state of OBE in the three outcome-based schools that participated in this study, brief, informal telephone interviews were conducted during July and August of 1994. In each case, school district personnel were very open and candid when responding to the interview questions. The following is a summary of information received during those interviews.

Two of the three outcome-based districts reported that their OBE initiative was continuing. The superintendent of the first district stated that they had been "remarkably unaffected" by the controversy in Iowa surrounding OBE. He stated that they sometimes even used the words "outcome-based education" when discussing school programs and issues. He surmised that the reason OBE had not become controversial in his district was because they had already identified the list of learner outcomes and defined them in their own unique terms. He named student assessment as the major challenge still facing the district, and reported that they hoped to meet that challenge through the use of technology.

The assistant superintendent of the second district related that OBE remained as the major school reform initiative and that "We feel good about our progress." The school board approved a policy which stated the student exit outcomes and related indicators for all grade levels. During the 1993-94 school year, the district piloted performance assessment
instruments for the exit outcome for grades 2, 5, 8, and 10. Scoring rubrics were developed for the performance tasks and nonschool personnel were trained to be performance assessors. The assistant superintendent attributed the success of OBE with community members to the conscious efforts by school personnel to make connections with leaders in business, the local community college, and parents. He predicted that by the fall of 1995, the district will have completed a training package that will include the student outcomes, indicators, assessment tasks, and scoring rubrics that will be used during comprehensive on-site training sessions for other districts.

The superintendent of the third district reported that they were no longer pursuing their involvement with OBE. As the controversy surrounding OBE was increasing in Iowa, he was asked by school board members "to keep the district out of the headlines." The superintendent reported that the words "outcome-based education" were considered "dirty words" and were no longer used in relation to the school program. Cooperative learning and site-based management were reported as the current reform initiatives being pursued by the district.


ACKNOWLEDGMENTS

Words are inadequate to express the gratitude I feel to so many people for their support and encouragement during my doctoral program. I first want to thank my family for their ongoing support of my educational and career adventures. My sons, Scott and Dave, not only agreed to be transplanted from our home in Germany to campus housing in Ames, but have continually encouraged and prodded me to follow my dreams. Together we actually thrived while sharing the cramped living conditions during Scott's senior year and Dave's freshman year of high school. My parents, Bill and Alice Ferguson, have always been sources of support and inspiration. They gave me the roots that allowed me to find my wings.

My appreciation is also extended to the members of my doctoral committee, Professors Dick Manatt, Larry Ebbers, Russell Mullen, Tony Netusil, and Shirley Stow. My major professor, Dick Manatt, coached me through every phase of my program and unselfishly shared his time and vast educational expertise. Other committee members also gave me ongoing support when I needed it and served as models of professionalism. I also want to thank Professors Barbara Licklider and Charles Railsback for encouraging my interest in and experimentation with outcome-based education. They provided many insights and rich opportunities for me to grow professionally.

I also want to thank my extended family, Marc Shellstrom, Judy Sessions, and Larry Sessions. Our study group time together helped me
grow academically and created some of my fondest memories of my time at Iowa State University. They remain cherished friends and colleagues.
TEACHER'S SELF-ANALYSIS
OF CLASSROOM ASSESSMENT PROCEDURES

This survey has been adapted from a survey
developed by Richard J. Stiggins

Richard J. Stiggins, Director
Center for Classroom Assessment
Northwest Regional Educational Laboratory
101 S.W. Main, Suite 500
Portland, Oregon 97204
DEMOGRAPHIC INFORMATION

Please complete the following demographic information.

What grade level are you presently teaching?

   ___ 1   ___ 2   ___ 3   ___ 4   ___ 5

How many years have you been teaching (please do not count the present school year)?

   ___ 0-1 year   ___ 2-5 years   ___ 6-12 years
   ___ 13-20 years   ___ more than 20 years

How many years have you been involved in outcome-based education (do not count this year)?

   ___ 0-1 year   ___ 2 years   ___ 3 years   ___ more than 3

How many years have you taught in this school (not counting this year)?

   ___ 0-1 year   ___ 2-3 years   ___ more than 3 years

What is your age?

   ___ 25 or under   ___ 26-35   ___ 36-45   ___ 46 or over

What is your gender?

   ___ female   ___ male

What is your education level?

   ___ BA/BS   ___ MA   ___ MA+30   ___ Doctorate

What is your racial background?

   ___ White   ___ Hispanic   ___ Black   ___ Asian
   ___ Multi-racial   ___ Other
PLEASE READ VERY CAREFULLY

Reason for the Analysis

Reading or Mathematics

This questionnaire has been designed to help you analyze your day to day assessments of student achievement. To help you analyze your assessment methods, we have devised a series of probing questions which will require you to think carefully about your testing procedures. This is not a quick questionnaire. It will take about 15 minutes of careful thought on your part. By combining your answers with those of many other teachers, we will be able to reach our goal of understanding the classroom assessment environment more clearly.

The ultimate value of this analysis depends on three key factors: 1) our success in clearly defining differences among several types of tests, 2) your success in clearly focusing on each different type of test when answering questions about it, and 3) your willingness to provide reflective, thoughtful responses. We are interested in learning about your use of and attitudes about six different types of achievement measures.

SPECIAL GUIDELINES FOR COMPLETING YOUR ANALYSIS

1. As you respond to the following questions, please distinguish carefully among the six types of assessments.

2. Many questions ask you to describe attitudes and practices you may not have thought about before. Please take the time to reflect on these questions before answering. This analysis is designed to provide indepth perspectives and may take up to 15 minutes to complete. Please set aside at least that much quiet time to complete it.

3. Our objective is to learn your actual assessment activities—not your practices as you think they should be. The self-analysis guide is completed anonymously to encourage you to reflect on and describe assessment as you really conduct it.

4. Several questions ask you to identify the percentage of your tests that have certain characteristics. Precise estimation will be difficult. Best guesses will suffice. Then mark the answer that includes your guess.

5. If any question fails to provide the response option you want, please mark the one that is closest to your response.

6. Finally, as you answer these questions reflect only on your assessment of student achievement in the subject you see circled at the top of this page, and the grade level you are presently teaching.
SECTION I - CLASSROOM DECISIONS

Tests can serve many purposes. That is, they can help you make any of a variety of decisions you face in the classroom. Before we ask about tests, however, we would like to better understand the decisions you face.

Listed below are several types of decisions. Please use the following scale to indicate the relative importance of each type of decision in your classroom. Remember, describe importance ONLY as it relates to the decisions you make in the subject area and grade you specified on the questionnaire cover.

SCALE:

  IMPORTANT  A  B  C  D  E  UNIMPORTANT

  1. Diagnosing the strengths and weaknesses of individual students.
  2. Diagnosing group needs.
  3. Grouping students for instruction.
  4. Determining the achievement potential of students.
  5. Assigning grades to students.
  6. Evaluating instructional unit to see if it worked.
  7. Communicating academic expectations.
  8. Controlling and motivating students.
SECTION II - TEST USE

Please use the following scale for the next section (questions 9-14) of the test to indicate the one that best describes your current level of use for each type of test:

SCALE:

A  I do not currently use them and do not plan to use them in the future.
B  I have decided to start using them in the future, but have not started to do so yet.
C  I currently use them, but I find they take great effort to use.
D  I use them on my own as a regular part of my instruction and do so comfortably.

___ 9. OBJECTIVE TEACHER-MADE PAPER AND PENCIL TESTS - this category includes all true/false, multiple choice, matching, fill-in and short answer tests and quizzes which you develop.

___ 10. TESTS EMBEDDED OR INCLUDED IN THE INSTRUCTIONAL MATERIALS - these may be found in an instructor's guide or may take the form of questions at the end of chapters in the materials themselves.

___ 11. STANDARDIZED ACHIEVEMENT TEST - these are offered by test publishers and also include state or district wide tests.

___ 12. ORAL QUESTIONING - these are used on a day to day basis during instruction to track whether individual students or the class as a group are learning the material.

___ 13. STRUCTURED PERFORMANCE ASSESSMENT - these include 1) a clearly defined reason for assessment; 2) planned exercises to elicit student responses; 3) a prespecified response to be evaluated; and 4) carefully spelled out scoring procedures.

___ 14. SPONTANEOUS PERFORMANCE ASSESSMENT - these include informal opportunities to observe and evaluate a student's performance and to judge the student's proficiency.
SECTION III - TESTING CONCERNS

Please use the following scale for the next section (questions 15-20) of the test to indicate the one that best describes your primary concern about the use of each type of test:

SCALE:

A I am concerned about my lack of information about using this type of test.
B I am concerned about my level of training, skill, and experience in using this type of test.
C I am concerned about the amount of time required to manage the use of such tests.
D I am concerned about how my students react when I administer this type of test.
E I am concerned about establishing working relationships with other teachers using this type of test.
F I am concerned about using these tests more effectively.
G I have no concerns about using this type of test.

___ 15. OBJECTIVE TEACHER-MADE PAPER AND PENCIL TESTS
___ 16. TESTS EMBEDDED OR INCLUDED IN THE INSTRUCTIONAL MATERIALS
___ 17. STANDARDIZED ACHIEVEMENT TEST
___ 18. ORAL QUESTIONING
___ 19. STRUCTURED PERFORMANCE ASSESSMENT
___ 20. SPONTANEOUS PERFORMANCE ASSESSMENT