INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

PREVIEW

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Computing Composite Scale Scores for Accountability: A Validation Study of Nebraska's District Evaluation Model

by

Chad W. Buckendahl

A Dissertation

Presented to the Faculty of the Graduate College at the

University of Nebraska in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Interdepartmental Area of
Major: Psychological and Cultural Studies

Under the Supervision of Professors James C. Impara and Barbara S. Plake

Lincoln, Nebraska

June, 2000

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 9976977

PREVIEW

UMI®

UMI Microform9976977
Copyright 2000 by Bell & Howell Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

DISSERTATION TITLE

Computing Composite Scale Scores for Accountability: A Validation Study of Nebraska's District Evaluation Model

BY

Chad W. Buckendahl, Ph.D.

SUPERVISORY COMMITTEE:

	APPROVED	DATE
Signature	James C Impara	June 21, 2000
Typed Name	James C. Impara, Ph.D.	
Signature	Barbara S Plake	June 21, 2000
Typed Name	Barbara S. Plake, Ph.D.	
Signature	Ellen Weissinger	June 21, 2000
Typed Name	Ellen M. Weissinger, Ph.D.	
Signature	Reece L. Peterson	June 21, 2000
Typed Name	Reece L. Peterson, Ph.D.	
Signature		
Typed Name		
Signature		
Typed Name		

GRADUATE COLLEGE
UNIVERSITY OF NEBRASKA

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Computing composite scale scores for accountability: A validation study of Nebraska's district evaluation model

Chad W. Buckendahl, Ph.D.

University of Nebraska, 2000

Advisers: James C. Impara and Barbara S. Plake

Many states use a statewide assessment strategy to evaluate districts on a common measure. Because districts in Nebraska are not measured on a common instrument, comparisons are much more difficult. This study examined an evaluation strategy for a state accountability model that considered school districts uniquely and classified them into an overall rating system (SPR) based on a combination (CSS) of salient factors. Classification decision consistency was analyzed to determine the appropriate model.

Student performance and non-cognitive indicator data for three grade levels and two content areas from the 67 school districts in the state of Florida were used. Data were also obtained from a Nebraska educational advisory committee. Regression and multiattribute utility theory (MAUT) methods were used to generate appropriate weights for model contributions. Analyses comparing classification decision consistency for three mathematical models were conducted using Spearman rank order correlations for composite scale scores (CSS) and kappa statistics for school performance ratings (SPR) classifications. Results show that there is a high level of agreement between the three mathematical models considered for the study indicating that the model that is easiest to understand and communicate should be recommended. The results lead to a discussion of the implications of the models for the Nebraska accountability system.

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Acknowledgments

I would like to acknowledge the contributions and support of individuals who made the completion of my doctoral program and this dissertation possible.

As advisors, mentors, and friends, James Impara and Barbara Plake were my constant companions on this journey. The different perspectives and experiences that they contributed to the process were invaluable as the line between theory and policy sometimes blurred. Committee members Ellen Weissinger and Reece Peterson were both extremely supportive, providing relevant feedback and flexibility in meetings.

Sarah Buckendahl, my wife and best friend has been incredibly patient with me through this process. I cannot thank her enough for being my biggest supporter and for answering the timeless question, "What does he actually do?" at family functions.

My parents, William and Sandra Buckendahl, always encouraged me to have an opinion and be independent. From reading to me as a toddler (even when it was the same story 4-5 times in a row) to our more recent discussions of society and theology, they have always tried to present the multiple perspectives that foster sound decision-making.

Bound initially by competition, Hoosiers, specifically Dan Endorf, Eli Moore, Jeff Moss, Darrell Stelling, and Jeff Wells, have been my models for personal and professional growth. Thanks to them for their support, friendship, and for keeping me around even when they found out that I was a defensive liability.

A special thanks is also extended to Douglas Christensen, Nebraska Commissioner of Education, for allowing me to use Nebraska's fledgling assessment and accountability system as an untested resource. It was a rare opportunity as an unknown to contribute to what may be a new direction in education policy in the country.

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE OF CONTENTS

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

List of Tables

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Computing Composite Scale Scores for Accountability: A Validation Study of Nebraska's District Evaluation Model

Chapter I

INTRODUCTION

Accountability Systems

Educational accountability is a common topic discussed among educators and administrators nationwide. It has become evident, though, that control over methods of accountability has shifted from the local jurisdiction (school districts) to the state jurisdiction (state departments of education and legislative entities). The shift is not surprising because popular media have given increasing attention to educational accountability. One reason may be the general belief that public education has not lived up to the expectations that have been placed on it.

The idea of educational accountability is not a recent phenomenon. In 1649, the Great and Central Courts of Massachusetts Bay Colony required that each town teach its children to read Scriptures (Marland, 1973). Any town that did not meet this requirement was fined five pounds. Later, England's Parliament commissioned the Newcastle Report in 1858 to conduct a comprehensive survey of English elementary education (Small, 1973). A goal of the study was to determine "what measures, if any, are required for the extension of sound and cheap elementary instruction" (p. 340). These examples represent early attempts at educational accountability.

More modern scholars have focused on outcome measures and the decisions that are generally associated with higher stakes accountability systems (Tyler, 1973; Dyer, Linn, & Patton, 1968). The validity information provided by those measures represents a

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

critical element in evaluating district performance. Whether using norm-referenced or criterion-referenced instruments, the alignment and proper use of those instruments relative to the desired educational objectives is essential to an accountability system.

In the last decade, researchers have continued to recommended accountability models that emphasize communicating meaningful results to constituents and stakeholders (Cornett & Gaines, 1997). One sentiment that they expressed was that it was necessary to explain it, if legislators were to trust it. An additional belief about accountability systems is that to be broadly supported, those who are being evaluated must have an opportunity to control some components of the evaluation (Law, 1999; King & Mathers, 1999). Allowing schools and districts to determine what is important in the curriculum helps to alleviate feelings of helplessness or apathy that may be associated with state mandates.

Additional problems arise when the scores are reported and schools' or districts' performance is compared. With a common assessment, states have rank ordered school districts based on performance at individual grades and content areas (e.g., Georgia) or on a composite index of district performance that considers non-cognitive[1] indicators (e.g., Kentucky) or one that does not (e.g., Texas). The rank ordering may not be meaningful, though, without also considering some of the non-cognitive indicators that may affect performance (Guskey & Kifer, 1990).

Evaluations of the efficacy of accountability systems have also been done to ensure that these systems address the needs of stakeholders that may be rewarded or sanctioned based on performance (Koretz, 1996; WMU Evaluation Center, 1995). These

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

evaluations examined the Kentucky Instructional Results Information System (KIRIS) as viewed by teachers and administrators. Although these evaluations of program efficacy were conducted by different organizations, both reports arrived at similar conclusions. Teachers and administrators felt pressure to perform at high levels (scores on the state assessment) as a result of the rewards and sanctions in the accountability system.

Based on an examination of the California accountability system, it was recommended that multiple assessments be used to measure performance and that subsequent reports be easy to understand for parents and policy makers (Association of California School Administrators, 1997). The report also concludes that standards-based accountability systems need to be clearly communicated to stakeholders for these systems to be supported.

State accountability systems generally use multiple components to measure district, student, and even teacher performance on common criteria. For most states, these accountability system components have included common assessment systems, student performance in those systems, and sometimes non-cognitive indicators used to adjust estimates of expected lower performance. The challenge for states is to combine these various components into a meaningful composite that meets accountability goals of measuring how schools or districts are performing that can be reported in an easy-to-understand manner. When high stakes are associated with the resulting composite variable, the challenge for equitable treatment of districts, while providing meaningful interpretations of performance is even more important.

PREVIEW

---

[1] Non-cognitive is defined here as demographic indicators that describe characteristics of students or school districts not related to cognitive ability. Examples of non-cognitive indicators include socioeconomic status, language barriers, or mobility. These indicators are generally considered uncontrollable by the students or school districts.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Purpose of the Study

The purpose of this study was to evaluate the utility of composite scores calculated for a state accountability model in terms of classification decisions using real data from school districts. The goal was to compare the utility of three increasingly complex model combinations with each other. Variables included in this study were as follows: 1) ratings of the technical quality of district assessments and 2) estimates of student performance from district assessments. A third variable that adjusted composite scores was included for districts that possessed student characteristics that presented high levels of challenge to district success.

This study was significant for two reasons. Empirically, it was important to examine the impact of model complexity on classification decision consistency for accountability systems that adjust school or district performance using district demographic characteristics. Consistency and stability of district classifications are essential for districts to trust the accountability system ratings. Politically, this study was important to inform Nebraska's policymakers whether their selected accountability model could effectively and meaningfully evaluate district performance without rank ordering districts on a single common measure. Because other state assessment and accountability systems rely primarily on common measures of student performance, the evidence of the accuracy and credibility of the accountability system was necessary to reinforce the soundness of Nebraska's decision to adopt the current assessment and accountability system.

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The Nebraska Accountability Model

As part of the Assessment and Accountability Plan adopted by the Nebraska Board of Education, school districts will be evaluated on how well they are meeting state adopted content standards. This study focused on the development of Composite Scale Scores (CSS) underlying the School Performance Ratings (SPR) that will be used to classify school districts on the desired evaluation criteria and to disseminate the results of the evaluation to appropriate stakeholders in a clear, understandable fashion. It was important that the components of the CSS and subsequent SPR classification system be related to the goals of the evaluation, contain components that were related to measuring student achievement, and provided all school districts an equivalent opportunity to perform well on the classification scale.

According to the Nebraska Department of Education, the SPR will initially contain three components with a fourth component added after the model has been in place for more than a year (Christensen, 1999). Each component was selected to be consistent with the adopted state assessment model and the relevance of each in measuring and explaining district performance on the state content standards. Another important consideration in the development of the CSS and SPR was a need for school districts to have equivalent opportunities to succeed within the proposed framework. This is why a rating system, rather than a ranking system, was proposed. If a numerical index that ranked districts was created, differences among those rankings may not be meaningful, yet could be erroneously interpreted as such. The following sections briefly describe the three components of the CSS and the subsequent SPR classification system.

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Component 1: Technical Quality of District Assessments

School districts will be required to provide documentation to an external review body that describes their overall assessment plan and contains information about the technical quality of their assessment strategies for measuring student performance on Nebraska's content standards (Nebraska Department of Education, 1999). These assessment strategies will measure student performance in Reading/Writing, Mathematics, Science, and Social Studies for grades 4, 8, and 11. The review body will use a variety of indicators and criteria for assessing the technical quality and adequacy of the district's assessment plan and instruments.

The review body will evaluate the districts' assessments and rate the technical quality of the assessments into one of five categories. A technical quality rubric developed by the Buros Center for Testing specifies the psychometric characteristics necessary to achieve a given technical quality classification (Plake, 2000). The technical quality of district assessments as evaluated by this rubric represents the first component in the SPR.

The rationale for including technical quality in the SPR is that it is necessary for districts to demonstrate the psychometric soundness of the methods they are using to determine student performance. Because the Nebraska Assessment and Accountability model does not rely on a single statewide assessment, each district will be responsible, in part, for the technical quality of the assessments they are using. If districts were only asked to provide student performance estimates without some assurance that the strategies or instruments they are using to measure performance meet technical standards, it would be difficult to interpret the results meaningfully.

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The technical quality criteria in the rubric represent characteristics of high quality measurement that are applicable in a school district (see for example, Traub, 1994; Anastasi, 1988; Ebel, 1979). Psychometric characteristics such as validity, reliability, opportunity to learn, fairness, and appropriate passing standards provide evidence of the technical quality of school districts' assessments. It follows, then, that higher technical quality assessments will produce student performance estimates that will likely be more credible than estimates produced from lower quality assessments.

Component 2: Student Performance on District Assessments

Districts will also provide information to the external review body describing student performance relative to the district's assessments. The specific strategies that districts use to determine the percentage of students meeting content standards will be provided in their assessment plans. Student performance described in a district assessment report (defined as the percentage of students meeting or exceeding a collective set of standards) represents the second component of the SPR.

The inclusion of student performance in an evaluation model is common. Currently, forty-one states use assessment scores of some type as part of their accountability system (Mather, 1999). The assessments from which these scores are procured range from norm-referenced (Florida Department of Education, 2000) to criterion-referenced (Connecticut Department of Education, 1999) to a combination of norm- and criterion-referenced assessments (Louisiana Department of Education, 1999). In contrast to other states, Nebraska's model is district-specific, meaning that it is possible for each district to have a unique combination of norm-referenced, criterion-referenced, and other assessments as part of its plan. Each district's plan may be unique.

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The technical quality of an assessment and students' performance on that assessment by themselves may not adequately represent a district's performance on the content standards. Therefore, additional factors that may affect district performance were considered in the Nebraska model.

Component 3: Non-Cognitive Indicators

The technical quality of districts' assessments and student performance relative to those assessments do not consider other non-cognitive indicators that may have a negative impact on students' academic performance. Acknowledging the potential effects of these indicators is admitting that school districts are not be able to control all factors related to its performance in the accountability model. To explain the potential impact these non-cognitive indicators may have on student performance, a scale value will be calculated. The indicators for the scale were selected following discussions with the Nebraska Department of Education and a review of relevant literature regarding non-cognitive indicators of educational performance (Mather, 1999; Reeves, 1998; Nelson, Yssledyke, & Thurlow, 1998; Jaeger, 1979; Land, 1975; Irvine, 1968; Cohen, 1968; Bauer, 1966). This scale represents the third component of the SPR.

The scale is normative, meaning that it can be used to describe the relative position of a district based on its proportion of students with specified non-cognitive indicators that may suppress scores on achievement tests. A majority of states that have incorporated non-cognitive educational indicators into their accountability systems have chosen ones such as parent(s)' level of education, dropout rate, expenditures, student behavior, learning or other disabilities, language barriers, mobility, and socioeconomic status (Mather, 1999). Because many of these indicators are difficult to measure,

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

researchers use variables that may serve as proxies for each. A proxy variable is one that is believed to represent the construct to be measured. For example, teenage birth rate has been used as a proxy for the emphasis a community places on education (Reeves, 1998). The proxy variables that were used in this model were as follows:

1. Percentage of students in free or reduced meal programs – proxy for socioeconomic status (SES).
2. Percentage of students with an Individual Education Plan (IEP) – proxy for learning or other disabilities.
3. Percent of students classified as Limited English Proficiency (LEP) or receiving English as Second Language (ESL) services – proxy for language acquisition barriers.
4. Ratio of average daily membership to enrollment – proxy for mobility.

Of the four non-cognitive indicators selected for the Nebraska accountability model, three were common to many states (King & Mathers, 1999; Cornett & Gaines, 1997). Only mobility was not a widely used variable nationally. Mobility was operationally defined here as a district's ratio of average daily membership to enrollment. Only four states currently collect data on student mobility (Mather, 1999). In an accountability system, the collection of student mobility information is important when districts are evaluated on their students' performance. If students have enrolled in a district, but are not currently on the roster, they have either dropped out or left the district. Conversely, new students who have not been taught in the school or district will not have had an opportunity to learn the content on which they are tested. This potential influx of students may have an adverse affect on a district's evaluation if these students

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

perform poorly because of the lack of control the district has over a student's learning opportunities[2].

The goal of the scale was to recognize and reward districts that have student populations that represent challenges to the district's ability to meet or exceed the content standards without penalizing districts that do not have similar student populations. Only districts that were rated above the state average in any of the identified challenge areas had the opportunity to receive adjustments on the CSS. This component of the SPR, then, was intended only to benefit districts with challenging populations by making an adjustment to their CSS and possibly the subsequent SPR. The intent was not to sanction districts that did not have a large number of challenging students. Because the scale is normative, though, some districts benefited from an adjustment from this scale whereas others did not. However, because districts were classified into rating categories, not rank ordered, districts within rating categories were not compared. All districts had equal opportunities to be classified in any rating category whether or not they received this adjustment.

Caution is urged when interpreting a scale that makes adjustments to an overall scale or rating based on uncontrollable variables. One problem when adjusting district ratings based on demographic characteristics is that such adjustments may confound interpretation of results because a district with lower test scores but a higher performance rating due to an adjustment may be classified above "satisfactory." "The students can't read any better because of the adjustment" (Irvine, 1999). Additionally, because the adjustment involves using variables that districts cannot control, adjusted scores should

[2] It should be noted that the proxy variable for mobility does not consider within district mobility. Because the evaluation system considers district performance, a district level proxy variable was used.

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

not be rank ordered because the scores do not have the same meaning for all districts (Reeves, 1998). As stated above, the Nebraska accountability plan uses a rating, not rank ordering procedure for evaluating the performance of districts.

Component 4: Improvement over Time

The fourth component of the SPR will be a district's improvement over time. The improvement component is sometimes called a district's Adequate Yearly Progress (AYP). In many states it is examined as a two or three year block as opposed to a single year. This is done because individual class differences from one year to the next may account for slight variations in a district's performance. By aggregating over time, the estimates that result from this analysis are likely to be more stable. At the outset of the SPR this component cannot be measured and therefore was not included in this dissertation.

Combining Components of the CSS

Over twenty years ago, Jaeger (1979) commented on how educational indicators were viewed by stakeholders. The sentiment expressed in his analogy may be even more appropriate today:

> "Many indicators are interpretable only because they have a known performance history; the Dow Jones Industrial Average is a case in point. Computation of the index is lost on the public, yet even novice investors and would-be investors respond to its level because it is prominently reported in most nightly news broadcasts and daily newspapers (p.299)."

If Jaeger's statement still holds true, the combination of these CSS components represents the most important consideration in the Nebraska model. Likely, this combination is the information that will be disseminated publicly. Because the CSS

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

involves three components measured on different scales, the various strategies used to create a composite variable of these unique pieces of information will probably yield different results. It is possible that various combinations of these components could then produce different classification decisions. To address this possibility, multiple combinations of the components were tested with school district data as part of a study to determine which model is most appropriate when considering interpretability (validity) and consistency (reliability) of classification. Three different combinations or models that differ in the complexity of how the CSS components are combined were tested. Three mathematical models were generated after consideration of literature on creating composite scores from three related areas: 1) an analogous practical situation (Kane, Kingsbury, Colton, & Estes, 1989; Schmeiser, 1987); 2) state determined accountability models (e.g., Florida Department of Education, 2000; Connecticut Department of Education, 1999; Norton, 1999); and 3) empirically tested accountability models (North Carolina Department of Education, 1999; Rothstein, 1999; Yap, 1997; Irvine, 1968).

PREVIEW

SPR Classification System

The School Performance Ratings (SPR) classification system consists of the composite scores generated from the mathematical models described above that are then transformed into five performance categories (NDE, 2000). A decision rule regarding the composite score ranges for each of these categories was created to specify the cut point for each level. The overall composite scale scores ranges from 1-100, however, for districts above the state average on non-cognitive indicator variables, it is possible for their scores to be as high as 110. The classification system score ranges are as follows:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

composite scores that ranged from 1 – 24 received a "1" classification[3]; composite scores from 25 – 40 received a "2" classification; composite scores that ranged from 41 – 60 received a "3" classification; composite scores that ranged from 61-75 received a "4" classification; and composite scores from 76+ received a "5" classification. It is not intended that the composite scale scores be disseminated because they may be misunderstood and used inappropriately as a method by which to rank order districts on this value. Rank ordering of districts would be inconsistent with the goal of the evaluation model that is to report performance on the standards.

Another important aspect of the SPR classification system is the labeling mechanism associated with each category. To emphasize the criterion-based nature of the SPR classification, language that does not connote a normative interpretation should be used. Such criterion-based language could be "Excellent," "Very Good," "Good," "Fair," and "Poor." Or, in keeping with the STARS (NDE, 1999) acronym, a number of stars (*) ranging from 1 to 5 could be used. This would eliminate classification language entirely, possibly avoiding potential misinterpretation. Other states (e.g., Alabama, Connecticut, Florida, Texas) also use classification systems to identify school, district, and even teacher (e.g., Tennessee) performance.

PREVIEW

Summary

The proposed SPR classification model has four strengths. First, it reduces potential misinterpretations associated with rank ordering school districts based on performance on a single indicator. Second, a rating system, as opposed to one that rank orders districts, aligns more with the Department of Education's goal of assisting districts with strategies for improving learning. Third, the SPR is analogous to the easily

[3] Numerical classifications are generic and will be replaced by language approved by the Commissioner of Education.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

interpretable Morningstar model used in the financial services industry (Harrell, 1997). In that model, financial products are given classifications (a number of "stars" ranging from 1 to 5) based in part on information on their technical quality and performance. A system such as this specifies the criteria for achieving ratings and financial products have equivalent opportunities in their peer group to achieve a given rating. Finally, using the SPR may encourage cooperation among districts to achieve the common goal of insuring that students are competent on the state content standards without fostering competition for ranked positions.

Many states utilize a single statewide assessment strategy (Mather, 1999). When districts are evaluated on a common measure comparisons across districts are easier. The Nebraska accountability model, though, emphasizes local control in the development of student performance measures. Because districts in the state are not being measured on a common instrument, comparisons are much more difficult. Researchers in this study examined an evaluation strategy for a state accountability model that considered each unit of analysis (school districts) uniquely and classified them into an overall rating system (SPR) based on a combination (CSS) of salient factors. Classification decision consistency was analyzed to determine the appropriate model.

The applicability of this study to the Nebraska accountability assessment and accountability system is clear. However, the results of the study are likely of interest to other schools, districts, or states that currently use indices, rank ordering or rating systems to evaluate measures of performance. Because many states rely primarily on common measures across all schools or districts, the ability for a district to select measures that they believe may be more aligned with their curriculum is hindered.

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Additionally, the credibility of an accountability system is damaged when it is viewed as unfair to those being evaluated or as a mandate from a state department of education.

An additional contribution of this study was the examination of recognized methods that have been used to generate weights for components in many state accountability systems. Although studies conducted for other states' accountability systems have used various methods, judgmental and empirical, to recommend component weights for composite scales or indices, application of these methods to an accountability system without a common measure of student performance was utilized in this study.

An essential component in an assessment and accountability system that encouraged districts to take an active role in the state's evaluation process by allowing them to choose the methods they use to assess students was examined in this study. By keeping the assessment methods more in the control of local districts, there is a greater chance that meaningful information for student learning could be used at the district level. At the state level, then, the rating system as opposed to a ranking system provides an estimate of how districts are performing on the content standards in a low stakes' environment. Given the consequences of low performance by schools or districts in many states, this may provide an alternative to other state-managed systems.

PREVIEW

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.