RAND EDUCATION

CHILDREN AND ADOLESCENTS
CIVIL JUSTICE
EDUCATION
ENERGY AND ENVIRONMENT
HEALTH AND HEALTH CARE
INTERNATIONAL AFFAIRS
POPULATION AND AGING
PUBLIC SAFETY
SCIENCE AND TECHNOLOGY
SUBSTANCE ABUSE
TERRORISM AND HOMELAND SECURITY
TRANSPORTATION AND INFRASTRUCTURE
U.S. NATIONAL SECURITY

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

Jump down to document

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

Purchase this document
Browse Books & Publications
Make a charitable contribution

For More Information

Visit RAND at www.rand.org
Explore RAND Education
View document details

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use.

This product is part of the RAND Corporation occasional paper series. RAND occasional papers may include an informed perspective on a timely policy issue, a discussion of new research methodologies, essays, a paper presented at a conference, a conference summary, or a summary of work in progress. All RAND occasional papers undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

OCCASIONAL
PAPER

External Audiences for Test-Based Accountability

The Perspectives of Journalists and Foundations

LAURA HAMILTON, BRIAN STECHER

OP-111-FF

March 2004

Prepared for the Ford Foundation

EDUCATION

This research described in the report was conducted by RAND Education for the Ford Foundation.

Library of Congress Cataloging-in-Publication Data

Hamilton, Laura S.
External audiences for test-based accountability : the perspectives of journalists and foundations / Laura Hamilton and Brian Stecher.
p. cm.
"OP-111."
Includes bibliographical references.
ISBN 0-8330-3579-7 (pbk.)
1. Educational accountability—United States. 2. Educational tests and measurements—United States. 3. Communication in education—United States. I. Stecher, Brian M. II.Title.

LB2806.22.H36 2004
379.1'58—dc22

2004004152

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

© Copyright 2004 RAND Corporation

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

Published 2004 by the RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
201 North Craig Street, Suite 202, Pittsburgh, PA 15213-1516
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

This paper examines the perspectives of journalists and foundation program officers on the accountability provisions of the No Child Left Behind Act of 2001 (NCLB) and offers recommendations for improving the presentation and communication of information about student outcomes. It complements other recent RAND publications on improving educational accountability. The results should interest policymakers and educators who are responsible for implementing NCLB, as well as journalists, foundation program officers, and others concerned about accountability in education. This study was supported by a grant from the Ford Foundation.

The RAND Corporation Quality Assurance Process

Peer review is an integral part of all RAND research projects. Prior to publication, this document, as with all documents in the RAND occasional paper series, was subject to a quality assurance process to ensure that the research meets several standards, including the following: The problem is well formulated; the research approach is well designed and well executed; the data and assumptions are sound; the findings are useful and advance knowledge; the implications and recommendations follow logically from the findings and are explained thoroughly; the documentation is accurate, understandable, cogent, and temperate in tone; the research demonstrates understanding of related previous studies; and the research is relevant, objective, independent, and balanced. Peer review is conducted by research professionals who were not members of the project team.

RAND routinely reviews and refines its quality assurance process and also conducts periodic external and internal reviews of the quality of its body of work. For additional details regarding the RAND quality assurance process, visit http://www.rand.org/standards/.

Contents

Preface iii
Figure ix
Acknowledgments xi

Introduction 1
Overview of the No Child Left Behind Act 4
Methods 6
Results 7
What Kinds of Test-Score Data Are Needed? 7
What Other Kinds of Information Are Needed? 8
What Problems Do Users Encounter When Interpreting Test Results? 8
Discussion 10
What Can States and Districts Do to Improve the Utility of Information for External Audiences? 10
What Can Researchers Do to Improve the Utility of information for External Audiences? 13
What Can Journalists Do to Enhance the Value of Information on School Performance? 15
What Can Program Officers Do to Enhance the Value of Information on School Performance? 17
Conclusions 19

Bibliography **21**

Figure

1. Elements of the No Child Left Behind Accountability Model 5

Acknowledgments

The authors express their gratitude to Eugene Maeroff and Richard Lee Colvin of the Hechinger Institute at Columbia University for allowing us to conduct our focus group in conjunction with their workshop and for facilitating our contacts with educational journalists. The authors are also grateful to Cyrus Driver of the Ford Foundation and David Tillipman of RAND for helping us identify appropriate foundation program officers to participate in our interviews.

Alice Wood of RAND and Abby Brown of the Pardee RAND Graduate School conducted the interviews. They also helped us summarize and interpret the responses, providing valuable input into the report. Brian Gill and Kerri Kerr of RAND reviewed the manuscript, and their suggestions greatly improved the final product.

Finally, the authors thank the educational journalists and foundation program officers who graciously participated in this study. We hope they find the results useful and thought-provoking.

Introduction

The No Child Left Behind Act (NCLB; P.L.107-110) has had a dramatic effect on the public education system since it was enacted in late 2001. NCLB requires states to adopt test-based accountability systems. These systems are designed to promote improved achievement by setting clear goals for student learning, measuring students' progress toward those goals, and holding educators accountable for this progress through the dissemination of results and the imposition of rewards, sanctions, and assistance. Although much has been written about the law, about states' reactions to it, and about its effects on schools, teachers, students, and parents, little has been written about the interaction of NCLB with external audiences that do not have formal connections with the schools. Yet these audiences, including the press and nonprofit organizations that promote educational programs and research, have roles to play in the success of NCLB. This paper looks at the information needs of two groups—journalists and foundation program officers—and at ways they might contribute to the effective operation of the NCLB accountability system.

These audiences are important because communication about performance plays such a key role in NCLB. The effectiveness of NCLB depends heavily on the actions taken by a variety of constituents in response to test-score data: Principals and teachers use information about performance to make instructional and organizational changes; districts use school-level information to identify schools that need assistance and to devise plans to help them; and parents use information about their children's performance, as well as the performance of their schools, to figure out how to assist their children and to make choices among schools or service providers. Effective actions on the part of any of these individuals will require clear reporting of results and thoughtful understanding of the meaning of school performance data.

The evidence suggests that we have not yet achieved either of these goals. For example, research indicates that users of test-score reports do not always understand or interpret them correctly (Koretz & Diebert, 1996; Hambleton & Slater, 1995). Furthermore, the reports themselves can seem contradictory, identifying schools as effective one year and "in need of improvement" the next (see, e.g., Gootman, 2003). Sometimes these apparent contradictions are due to changes in emphasis from reporting overall averages to reporting separately on subgroups of students. In so doing, the new rules may be revealing inequities that schools have managed to hide until now and increasing the likelihood that parents and others will put pressure on those schools to improve the performance of all students (Hall, Wiener, & Carey, 2003). Whatever the case, by tying actions to outcomes, NCLB has contributed to a transformation of the public conversation about education from one concerned

with equality of inputs to one focused to a large extent on the outcomes that schools produce. As a result, a better understanding of the use of information on student and school performance is essential for effective implementation of the law.

In an earlier study of the NCLB approach to accountability, we reviewed the literature on educational accountability systems and convened a discussion forum of 15 scholars who had studied accountability from a variety of perspectives. The group included researchers as well as practitioners representing district staff, school administrators, and teachers. The conversations with these experts, along with some review of the literature on test-based accountability, contributed to a RAND Corporation White Paper (Stecher, Hamilton, & Gonzalez, 2003) that describes the conditions necessary to make NCLB accountability work and provides guidance for those responsible for designing and implementing NCLB-compliant accountability systems.

Our earlier study revealed a number of ways in which existing accountability systems failed to meet the needs of educators, parents, and other stakeholder groups. We generated guidelines for improvement that related to each of the major components of the NCLB accountability model: content standards, assessments, achievement standards, adequate yearly progress, incentives, reporting, technical assistance, and parental choice (including the choice both of schools and of supplemental services). For example, experts at our discussion forum noted the importance of content standards that are written to facilitate curriculum design and instructional planning and of ensuring stakeholder access to the full content standards. On the topic of incentives, forum participants expressed concern about inappropriate behaviors that may result from high-stakes accountability and recommended ways in which incentive systems might be designed to encourage effective responses on the part of educators as well as students.

The present study focuses on two key external audiences that are involved in public education and that frequently use data from schools and districts in their work—print journalists and foundation program officers. The earlier study provided some guidance for making test-score data more useful for educators and parents, and our questions to journalists and program officers were informed by those findings. A number of other external audiences could have been included in this research—for example, community-based organizations and advocacy groups. We chose instead to focus on journalists and program officers in large part because the sponsor of the study expressed interest in their points of view and because we interact frequently with these groups in our role as researchers. In addition, journalists play a critical role in informing the public about how schools are performing and about what options are available to dissatisfied parents. Journalists also shape public opinion through the stories they write, and high-quality information about schools' performance is essential for accurate reporting.

Foundation program officers who work in the area of education also use information about school performance in their work. Many foundations operate programs that provide assistance and guidance to struggling schools, and those responsible for designing and overseeing those programs need accurate information about school operation and performance. Through their funding decisions, foundations can exert a strong influence on the kinds of programs and policies that are adopted in schools and on the type and quality of education research that is conducted.

Thus, through their daily work, foundation program officers and journalists can shape public education as well as public opinion about the education system, and both groups need valid and reliable information to do their work effectively. Moreover, members of both groups have experience communicating with parents and other members of the public, and their views may be useful for states or districts as they make decisions about implementation. Although many of the needs of these two groups may be different from those of other users of accountability data, much of the information they provided to us could help improve the utility of accountability systems for other stakeholders.

In this paper we use insights from conversations with journalists and program officers, in combination with the advice we received from participants in our earlier discussion forum, to identify additional ways to improve the utility of information from test-based accountability systems. We begin with a brief description of the main components of NCLB. Then we describe the methods we used to gather information from journalists and foundation program offices and summarize what we learned about their data needs and the problems they encounter when trying to interpret test results. Finally, we synthesize information from those conversations and from our earlier analysis to offer advice to the producers of performance reports and analyses (states, school districts, and researchers) and to the external consumers of those reports (journalists and program officers) on how to enhance the utility of the information.

Overview of the No Child Left Behind Act

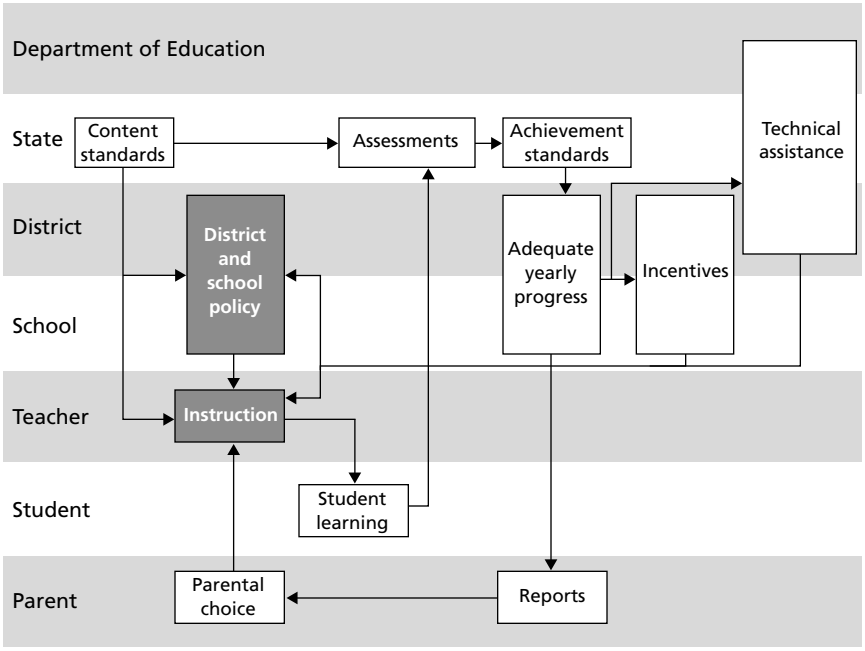
NCLB mandates that each state enact a system that includes the accountability components illustrated in Figure 1. The components of the system are indicated by the labeled boxes, and the arrows connecting the boxes indicate the flow of information, responsibility, or consequences. The horizontal bands in the figure indicate which level of the educational system—from the federal Department of Education down to local schools, students, and parents—is responsible for, or is the recipient of, each component. Boxes that extend across more than one band indicate multiple levels of responsibility.

The figure illustrates the main components of NCLB as well as the roles of key constituents, including state departments of education (SDEs), districts (local education agencies, or LEAs), and schools. States are required to adopt content standards in reading, mathematics, and science; these standards are intended to guide decisions about policy and instruction at the district and school levels. The content standards serve as the basis for the development of tests that measure student progress. In addition, states must establish performance standards that convert test scores into judgments about proficiency. These take the form of *cut scores* on the state tests—scores that indicate what level of performance will be classified as (for example) basic, proficient, or advanced. NCLB requires that all students reach the "proficient" performance level within 12 years, and the state must set increasing annual targets to ensure that this occurs. These targets are used to determine whether each school, district, and state has made "adequate yearly progress" (AYP).

Incentives, or systems of rewards and sanctions, are based on whether schools have met their AYP targets. Schools that meet the AYP targets can be rewarded; those that fail to meet the target for two years are identified as "needing improvement." Being classified as needing improvement triggers sanctions, including offering parents the choice of transferring their child to another school and offering students supplemental educational services (e.g., tutoring) from outside providers. Schools that continue to underperform for three or more years face stronger measures, including possible restructuring and eventual takeover.

The law does not prescribe specific actions for districts and schools to take to improve student performance. It places a strong emphasis on scientifically based practices but does not mandate any particular approach. So administrators' leadership skills and teachers' instructional skills will determine whether schools continue to improve and are successful, even though neither is a formal component of NCLB. This is shown in the figure by the shading of the boxes labeled policy and instruction—to a large extent, NCLB gives local educators control over these decisions.

Figure 1
Elements of the No Child Left Behind Accountability Model



RAND *OP111-1*

The complexity of the figure illustrates the roles played by a variety of stakeholders and suggests that the ways in which any of these groups respond and the conditions surrounding their responses may influence the degree to which a state's accountability system exerts positive or negative effects. In our earlier study, we gathered input from a variety of experts involved in educational accountability; many of their recommendations were intended to promote the provision of test-score information that serves the needs of educators and parents. By seeking additional input from journalists and foundation program officers, we hoped to refine our understanding of what is necessary to make test-based accountability contribute to improved public understanding and effective decisionmaking and to explore the needs of other stakeholder groups that were not the focus of the earlier work. To do this, we needed to understand how these groups used test-score information and what kinds of problems they encountered with it.

Methods

We conducted interviews with journalists and foundation program officers during the spring of 2003. For journalists, we took advantage of a workshop on testing sponsored by the Hechinger Institute on Education and the Media at which a number of education journalists from around the country were gathered. The workshop focused on reporting on educational testing, so all of the participating journalists were involved to some extent in writing about testing and accountability. We sent letters to all participating journalists, inviting them to share their perspectives on accountability, and six of the 24 workshop participants agreed to meet with us. We interviewed them in a focus group setting lasting about two hours.

Following the journalist focus group, we contacted approximately two dozen foundation program officers who were involved in projects related to educational accountability. Thirteen agreed to participate in individual telephone interviews with us. The telephone interviews lasted about an hour each. We used the same general set of questions to interview both the journalists and foundation program officers. The questions concerned participants' use of information from states and from researchers on student performance, their desire for additional information on student outcomes, the problems they encountered using the information, and their overall views on testing and accountability. We also asked them to share ideas about resources that would help them do their jobs better and to identify ways in which the producers of the information could be more helpful than they had been in the past. We ended the interviews by inviting them to provide any additional comments that they did not have a chance to share during the formal part of the interview.

The views of the program officers and journalists who participated in this study should not be construed as representing those of all program officers or journalists. There are, in fact, reasons to believe that our group of interviewees may differ from other members of these professions. The journalists who applied to, and were selected for, the Hechinger workshop had displayed interest and accomplishment in writing about education and may be unusually sophisticated with respect to education data. Similarly, the program officers we interviewed had all attained some level of expertise in education and were interested in accountability, so they are likely to be more knowledgeable about the topic than other program officers who may work with education data but who do not specialize in that area. These somewhat unique features of our participants should be kept in mind when interpreting the summary of their responses.

Results

Both groups had much to say about how states and researchers could do a better job of communicating performance data more effectively. In many cases, journalists and program officers provided similar responses, so much of the discussion below pertains to both groups. Instances in which the responses were different are noted.

What Kinds of Test-Score Data Are Needed?

Both groups found school-level performance reports useful. In particular, they appreciated the disaggregation of scores by student group that is now required under NCLB. One journalist noted that the achievement gap among racial/ethnic groups had been a frequent topic of news articles, so having access to data on the size and nature of the gap was helpful. Program officers used school-level data for a number of purposes—including selecting grantees, evaluating the effectiveness of programs, and engaging in advocacy—and many said they appreciated the availability of such data on state Web sites.

In addition, both groups demonstrated a relatively sophisticated understanding of state test-score data and recognized many of the limitations inherent in school-average scores.[1] They offered a number of suggestions for ways to improve the utility of data, most of which involved the creation of methods for linking student data over time and linking student data to teacher information. Several of the program officers were familiar with recent work on value-added modeling (one had recently funded such work) and thought this approach not only provided some important information about teacher quality but also represented the state of the art in approaches for evaluating test-score information. In addition, respondents were aware that student mobility rates, which are high in some districts and schools, limit the validity of inferences about performance changes at the school level. Several foundation program officers suggested that a student identifier system to track students as they change schools and link students to their teachers would help provide better information. Similarly, respondents in both groups expressed a desire to follow students as they leave the K–12 system and enter higher education or the workforce and noted the potential value of an ID system that would facilitate such tracking of students. Respondents were aware of the political and technological challenges associated with this type of data system, particularly

[1] As noted earlier, the journalists we interviewed may be unusually sophisticated in their understanding of this topic and are unlikely to be representative of all journalists.

the student-teacher link, but they also strongly believed it would help them do their jobs better.

What Other Kinds of Information Are Needed?

A majority of participants from both groups expressed a desire for additional information beyond test-score data. Although they recognized the utility of test scores, many respondents said that student work samples would provide a much more complete picture of students' capabilities than is evident in test scores alone. Journalists thought that being able to publish samples of student work, such as essays, would make their stories more vivid and help the public understand what students can and cannot do. Program officers said that student work samples would give them a better understanding of what is occurring in the classroom and allow them to evaluate their programs more effectively. One mentioned a presentation in which he showed other program officers a series of slides depicting a single student's work as it evolved over a ten-week period and noted that the presentation left a strong impression on the audience.

Participants were also interested in contextual information about schools, classrooms, and students' families. Specific items that members of both groups mentioned included teacher qualifications and turnover rates; class size; school funding and other resources; and student background data, such as poverty levels. Some respondents noted that school and student background information is critical for interpreting test scores; one program officer commented that it is difficult to know whether a score at the 35th percentile is good or bad without knowing something about the conditions of the student's home and school. A small number of program officers also stated that data on instructional practices, relationships between teachers and students, and other classroom characteristics that are difficult to measure would be helpful if there were a way to obtain them.

Finally, program officers in particular wanted better information on graduation and absentee rates. A number of them expressed frustration with the poor quality of measurement of graduation and absentee rates and with the differences in measurement methodologies used across districts and states.

What Problems Do Users Encounter When Interpreting Test Results?

Both groups mentioned a number of factors that hindered their effective use of test-score data, although the primary concerns of the two groups were somewhat different. For journalists, the timeliness of reporting and the completeness of the data given to them were the most frequent concerns. Every one of the participating journalists stated that the need to produce stories on deadline required them to obtain and analyze data very rapidly. Some states and districts provided journalists with partial data in advance of the public release of scores, but these data were perceived as insufficient for producing complete and accurate stories. For example, in one state the school-average scores were released a few days before the disaggregated subgroup data, forcing journalists to decide whether to omit the latter from their articles or to try to convince their editors to publish a follow-up story. Journalists also

stated that they often received results without the necessary accompanying explanations of how the results were generated or what they meant. Newspapers are sometimes forced to publish corrections a few days after the initial data release because journalists lack the time and information to do the necessary quality control before the results are released.

Another concern specific to journalists was their inability to obtain what they believed to be objective information. They noted that press releases often carried the district's or state's "spin" on the data and that access to officials who could provide more information was sometimes limited. For example, one journalist stated that district personnel often attributed score changes to particular programs or initiatives but failed to provide evidence to back up these assertions; she noted that journalists typically have no way to determine whether this attribution is justified. Another said that she was rarely given the opportunity to interview the superintendent: During telephone press conferences, the phone lines operated in only one direction, so the journalists could hear but could not ask questions. All our participants said they needed better access and more objective information.

Journalists also raised concerns that were related to their roles in the organizations for which they work. They reported that editors sometimes prevented them from writing the stories they thought were important because editors, like the state and district administrators providing the data, often wanted to shape the story in a certain way. Participants said that even when reporters maintained control over the content of the story, editors often shaped the message through the headlines they attached to the stories. A few also said that the public demands certain types of stories, making it difficult for journalists to report some of the issues that they believe are important. They complained that both editors and members of the public often exerted pressure to simplify, when in fact the real story is complex and should be reported that way.

Program officers' concerns focused more on the validity of information from testing systems and on their need for more sophisticated analysis than what states and districts typically conduct and publish. A specific issue mentioned by several participants was the difficulty of comparing results across states that use different tests, have different standards, and apply different methodologies and reporting strategies. Even within a state, some program officers said it was often difficult to compare results over time or across districts because of variations in reporting strategies.

Both groups expressed some frustration with their inability to understand the methods used to create scores. Some attributed the problem in part to their own lack of analytical training, but they also noted that states and districts typically failed to provide methodological details when they released scores. In particular, there was interest in a greater level of detail on how performance levels (or cut scores) were set, in order to facilitate interpretation of what the levels mean. A related issue was a lack of evidence to convince users of the test data that the results are believable and not the result of factors such as inappropriate test preparation or data manipulation.

Discussion

The feedback we received from journalists and foundation program officers suggests that some changes to the way test scores are reported would facilitate greater understanding. By combining the insights from these interviews with the information we gathered in our earlier study and in other RAND work (Stecher, Hamilton, & Gonzalez, 2003; Hamilton & Stecher, 2002), we developed a set of guidelines that are intended to help not only journalists and program officers but also other users of large-scale test data. Throughout this discussion, we use the interview results to illustrate the need for specific actions, but the guidelines represent our synthesis of lessons learned from a number of sources including the interviews. They reflect our opinions, informed by these various sources. We begin with suggestions that are intended for state and district personnel who design and implement accountability systems and for researchers who study and report on them. We then turn to some specific guidance for journalists and program officers.

What Can States and Districts Do to Improve the Utility of Information for External Audiences?

States and districts have the primary responsibility for communicating test-score results and other information about school performance. There is much they can do to make the data more useful for these key external audiences. We recognize that districts and states currently face a number of demands related to accountability under NCLB and that they may lack the necessary time and resources to follow all these suggestions. In particular, many districts and some smaller states lack the personnel as well as the infrastructure needed to develop new data and reporting systems and must collaborate with outside contractors or other organizations to comply with NCLB's reporting requirements. Nonetheless, we believe that the actions described below, if implemented, could ultimately benefit states and districts and are worth considering as they move forward with their plans to generate and share data.

Provide data in a format that facilitates independent analysis.

Journalists, in particular, noted that they often receive scores in portable document format (PDF) or in hard-copy form, which made it difficult for them to conduct any additional analysis of the data. They stated that they would prefer a file that could be analyzed, such as a spreadsheet or database file. In addition, journalists stated that the results they received were sometimes combined in ways that made them difficult to interpret. For example, one par-

ticipant stated that her district collapsed all score levels into two categories—proficient or not proficient—rather than providing information about all four categories of performance. In her opinion, this practice results in an unnecessary loss of information. States and districts should try to make the information more useful by providing it in an analyzable form that includes as much detail as possible without compromising student confidentiality. State and district officials should then make themselves available for follow-up questions from the media as well as from educators and parents.

Publish data in a timely manner.

Journalists expressed concerns about obtaining access to data as early as possible and about the accuracy and completeness of those data. To a lesser degree, some of the program officers also mentioned this as a concern. Although states and districts clearly must engage in quality-control procedures that will delay the release of results to some extent, the public is ultimately best served by having access to data earlier rather than later. Teachers and principals need the information to make changes to their instructional programs; parents need it to inform choices about schools and supplemental services; journalists need it to respond to pressures for timely reporting; and program officers need it to evaluate their programs and provide assistance to struggling schools and districts.

Provide clear explanations of methodology.

Reports of test-score data should be accompanied by clear, accessible information to help users understand how the scores were generated and what they mean. To the extent possible, this information should be presented in a concise form but should address important technical concerns, such as the reliability of test scores and the approaches used to ensure that the test is not biased against particular groups of students. One of the most frequently mentioned methodological concerns was the need to understand how cut scores (e.g., between proficient and not proficient) were set. One of the presumed benefits of reporting in terms of performance levels or cut scores is that it conveys more meaning than is inherent in the use of ranks or other frequently used reporting strategies. As noted earlier, however, there is evidence that educators, policymakers, and members of the media often have difficulty interpreting those scores (Koretz & Diebert, 1996; Hambleton & Slater, 1995). One problem is that the descriptions that accompany the labels—such as "proficient" or "basic"—are often vague, as some of our respondents noted. Similarly, our respondents said that they needed additional information to interpret these performance levels. Reports of results should not only provide descriptors of what kinds of skills and knowledge are represented by each level but should also explain how the levels were set and how much measurement error is associated with them. These explanations would be helpful to all users of the data—especially parents, who need to understand what the scores mean in order to determine how best to help their children.

Supplement test-score data with qualitative data on student performance and data on school inputs.

Qualitative information can inform the interpretation of quantitative test score data by providing context. For example, districts could produce a bank of student work samples that illustrate what students are doing at various grade levels and that provide some indication of

the difficulty level of the work. Work samples not only would enhance public understanding of what schools and students are doing but could be especially helpful for parents who need to understand the nature and difficulty of the work expected of students and who want to work with their children to improve their individual performance. Of course, student work samples and other sources of qualitative data are not necessarily straightforward to interpret. If used, work samples could serve as exemplars to illustrate the kinds of products that are expected of students and the kinds of problems students encounter in school, but they should be only one source of information among many.

Data on school inputs, such as resources and teacher qualifications, were also of interest to our interviewees, who reported having difficulty getting access to such information. Such data are not always easy to provide, and additional work needs to be done to identify the kinds of information that would meet the needs of journalists, program officers, and other users of the data. For many inputs of interest there are few, if any, valid measures, so work is needed to develop methods for assessing and reporting on these inputs. As we discuss below, this is one area in which researchers could make a contribution.

Provide examples of test items and correct responses.

Released test items and samples of student responses to open-ended questions (e.g., essays) serve a similar purpose as the qualitative information described above. In addition, this information may help build understanding of, and support for, the accountability system by showing stakeholders how test performance is linked to standards. By showing sample items and responses linked to descriptions of the standards they measure, states could provide concrete evidence of how the test captures the standards and could thus help educators, parents, and other members of the public better understand the role that standards play in the accountability system.

Create systems to monitor curriculum and instruction.

Interview participants expressed concerns about the validity of scores from accountability tests and about possible negative effects of testing on curriculum and instruction. NCLB's testing requirements promote a focus on reading and mathematics (and, to a lesser extent, science). Although no provisions of the law explicitly call for de-emphasizing other subjects, and in fact the law includes language regarding the importance of subjects such as the arts,[2] past experience with test-based accountability systems suggests that teachers are likely to reallocate their time from untested to tested subjects (Jones et al., 1999; Shepard & Dougherty, 1991; Smith et al., 1991; Stecher & Barron, 1999; Stecher et al., 2000). Reallocation also occurs within subjects; high-stakes testing has led teachers to emphasize particular formats (such as multiple choice), topics, and skills over others (Shepard & Dougherty, 1991; Romberg, Zarinia, & Williams, 1989; Smith et al., 1991). Although some reallocation may be warranted, it is important for administrators and policymakers to monitor it and to help teachers find ways to maintain an appropriate balance across subjects, topics, and skills so that students continue to receive a balanced curriculum.

---

[2] For example, the law includes as one of its objectives: "to support systemic education reform by strengthening arts education as an integral part of the elementary school and secondary school curriculum" (NCLB, 2001).

Design testing systems to minimize score inflation.

Excessive or inappropriate test preparation can lead to an increase in test scores that is not accompanied by a corresponding increase in student achievement, a phenomenon known as *score inflation* (Koretz, 2003). Inflated test scores lose their utility for helping teachers, parents, policymakers, and others make effective decisions about how to improve performance. Testing systems can be designed to minimize the likelihood of score inflation. Some approaches include changing the test items each year and varying the formats of the items rather than relying on a single format such as multiple choice. These steps reduce the likelihood that teachers will focus on a single set of items or item type as they prepare students for the test. Professional development activities that help teachers understand what kinds of test-preparation practices are appropriate and effective would also help to prevent score inflation, particularly if teachers are convinced that they can raise scores without engaging in excessive teaching to the test. In addition, the state can use the results of National Assessment of Educational Progress (NAEP) or other auditing mechanisms to identify score inflation in its early stages so that additional steps can be taken to reduce it. Additional suggestions for addressing score inflation are provided in Hamilton and Stecher (2002).

What Can Researchers Do to Improve the Utility of Information for External Audiences?

Researchers also influence the utility of information from test-based accountability systems. Frequently, they are the ones who analyze the raw data provided by states and districts. Certainly, they are the group that is typically tasked with developing new measurement and reporting systems. Thus, our recommendations for researchers focus on researchers' roles in both analysis and reporting.

Improve systems for measuring non-test student outcomes and school inputs.

One of the most frequently mentioned sources of frustration in our interviews was a lack of accurate measures of outcomes other than state test scores, and of inputs such as financing and curriculum. Several program officers, in particular, mentioned the inconsistent measurement of graduation rates across states as a problem for their use of data. Researchers have recently played an important role in documenting some of the technical problems involved in measuring graduation rates and providing states and districts with advice on how to improve it (Greene & Forster, 2003; Swanson, 2003), and they could contribute to improved measurement of other important outcomes as well as inputs. For example, accurate information on students' course-taking histories, schools' academic environments and disciplinary actions, and teachers' practices could help users of test-score data obtain a richer and more balanced picture of what schools are doing and could be particularly helpful to parents who have to choose among schools or who want to understand what factors may contribute to their schools' performance. Researchers should work closely with state and district personnel on these activities to ensure mutual understanding and to increase the likelihood that researchers' ideas will be tested in real school settings and, if shown to be valid and effective, that they will be implemented in practice.

Conduct studies to compare state systems.

Several respondents requested more information on differences among states' accountability systems. Users of accountability data often have the need, or at least the desire, to compare one state's practices or outcomes with those of other states. To do so, they need accurate, accessible information about differences in the features of state systems—e.g., the nature of states' testing programs and the difficulty of their performance standards. Journalists were especially interested in this information as a way to place local findings in a broader context. Both groups thought it would be valuable to compare states' methodologies for reporting data on graduation rates and other non-test outcomes. Although participants acknowledged that some comparative data are available, such as those compiled each year by *Education Week,* they were interested in more detail, particularly on the technical characteristics of state assessment and reporting systems.

Synthesize findings into a coherent story and distribute widely.

Journalists have noted that research results are often difficult to use, in large part because of the multitude of conflicting studies and because research reports are not written in a clear and understandable format (see also Colvin, 2002/2003, for results of a survey of journalists that addressed this issue). Some of the journalists who participated in our study suggested that researchers could provide a valuable service by producing syntheses that contain short, accessible summaries of what is known about the effectiveness of specific programs or policies. Although existing research results may not always lend themselves to clear bottom-line conclusions, summaries that provide straightforward and concise descriptions of existing research and, where possible, some analysis of what kinds of conclusions are warranted could be useful not only to journalists but also to policymakers and practitioners.

Report research in a timely, concise, and clear manner.

Several interview participants expressed a desire for research results that are presented in "plain English." In addition, they commented that existing results are often provided long after a program has been implemented and are too late to affect decisionmaking. Program officers expressed a strong desire for timelier reporting of research results, perhaps even before a study is completed. They urged researchers to find ways to share interim results and to make them accessible to a nontechnical audience. In addition, members of both groups noted that providing results in multiple formats, including graphs or other pictorial aids, could make results more easily interpretable to a wider audience.

Collaborate with practitioners and the media.

Several program officers indicated dissatisfaction with the extent to which research findings made their way into schools and classrooms. One reason for this failure of research to be translated into practice may be the generally low level of collaboration between researchers and practitioners, particularly teachers. By working directly with practitioners, researchers will not only increase the chances that important findings will influence what happens in schools but will also be more likely to conduct work that meets the needs of practitioners, thereby improving the quality of researchers' work. Researchers should also collaborate with the media to help them identify important stories in the data and to improve the dissemination and interpretability of results. This advice is relevant to all areas of education research,

but it may be particularly important for helping schools and districts adopt better accountability systems and meet the federal law's demand for scientifically based research.

Provide seminars and expert help when needed.

Researchers should make themselves accessible to journalists, program officers, and other users of data on an as-needed basis. Researchers could also enhance users' understanding of data by holding seminars for targeted audiences. Journalists, for example, said they would appreciate seminars not only for reporters but also for editors, since editors often add an interpretive layer to news stories. They thought it would be helpful for researchers to provide advice and guidance designed to avoid oversimplification and to explain key technical issues, such as margin of error or methods for setting cut scores. Seminars or other resources (e.g., short briefing papers) on "how tests work"—how they differ from one another and change over time, what factors affect their difficulty, and how users should interpret score changes—were seen as especially beneficial.

Gather information on the needs of other external audiences.

Data collection for this study was limited to two groups, but a number of other users of accountability data should be consulted on a periodic basis to assess whether data from accountability systems are meeting their needs. These include political and community-based organizations and legislators. Researchers could follow up on our study by gathering information from these other groups and from other types of journalists and program officers (e.g., those who do not work primarily in education but who occasionally use data on schools).

What Can Journalists Do to Enhance the Value of Information on School Performance?

Journalists have an important role to play in the NCLB accountability system. To a large extent, it is through the media that parents and community members find out about school performance. Journalists may not be aware of the key role they play in NCLB, but they act as filters on the performance information provided by states and districts. The manner in which they report the information and the stories they tell influence parents' attitudes and actions. With this in mind, we suggest that journalists consider the following types of activities.

Press for information on non-tested outcomes.

The journalists who participated in our study indicated that they would like access to information other than scores on state accountability tests. Many of the program officers echoed this view. As journalists conduct research for their stories, they should make it clear that they need additional information such as student work samples, so that those responsible for providing data are made aware of the importance of this information. To understand how accountability is affecting schools, journalists should gather evidence on non-tested topics and subjects—for example, whether instruction in other subjects such as social studies is being shortchanged. Another important source of information may be found in national comparison data that help readers place their state's results in a broader context. If necessary, jour-

nalists should seek out the assistance of researchers and others when attempting to gather and interpret this information.

Pay attention to the specific reasons that schools are labeled as needing improvement.
Journalists should try to obtain the details behind the labels—e.g., What characteristics are common among schools that fail to make AYP and how do these schools differ from one another? For a particular school that has been designated as needing improvement, is the designation a result of widespread failure of students to meet the test-score targets, or does it stem from an isolated problem, such as not meeting the requirement that 95 percent of all students take the test? By providing readers with a story that goes beyond which schools have met their targets and which have not, journalists will contribute to an improved public debate about testing and accountability and about public education more generally.

Use teachers, students, and parents as sources of information.
Understanding what is happening in schools labeled as needing improvement will require access to teachers and classrooms. More generally, the public's understanding of their state's testing and accountability program would be enhanced by information on school and classroom practices, including the extent of curriculum narrowing (if any) and test preparation practices. In addition, it is important to explore the effects of accountability on such nonacademic outcomes as teacher and student morale. Finding out, for example, whether teachers perceive resources to be sufficient and whether they perceive the state's expectations for performance gains as realistic could provide an important angle to stories about accountability. Moreover, talking with teachers is likely to give a different perspective on test-score results than what is typically provided by the superintendents and other administrators who develop or contribute to press releases.

Parents represent another important group that should be included in journalists' investigations. Conversations with parents could help journalists understand how parents interpret test-score data, whether they perceive the score reports to be easy to use, what they think the results signify about their children and their schools, and whether parents plan to take any action in response to results. Parent empowerment is one of the goals expressed by many advocates of test-based accountability, and there is a need for information on the extent to which this goal is being achieved. Journalists can contribute by gathering information directly from parents.

Work with researchers to improve the dissemination and interpretability of research findings.
As we discussed above, journalists and program officers noted that findings are not always presented in an easy-to-understand format. Timeliness of reporting is a problem as well. By making themselves available to researchers for consultation and collaboration, journalists can help researchers make their results more accessible to practitioners and the public and can help researchers disseminate their findings more effectively.

What Can Program Officers Do to Enhance the Value of Information on School Performance?

Foundation program officers play a distinctive role in the educational improvement process. They fund innovation, and they use data to make judgments about the success of their efforts. As consumers of test results and as developers of new educational activities, program officers are, in fact, part of the larger NCLB-related school improvement enterprise. We suggest that foundation program officers adopt the following strategies.

Require programs to collect data that go beyond state test scores.

Program officers could improve their own understanding of the effects of programs and interventions and could help shape the public debate on educational outcomes by requiring grant recipients to collect and report data other than scores on state accountability tests. Although such scores are perceived to be useful and should continue to be analyzed, additional information would provide a richer description of schools and classrooms. This information could include, for example, scores on tests in subjects that are not part of the accountability system, student work samples or other classroom artifacts, or information on classroom practices and student-teacher interactions. Again, it is important to recognize that methods for measuring some of these outcomes and inputs are not well developed, so all such information should be seen as part of a broader array of information on school practices and student performance.

Consider funding initiatives to improve data systems and develop new measures.

Much of the information that program officers told us they were lacking could be obtained through improved state or district data systems and through new measures of outcomes that are currently not included in the accountability system, such as graduation rates. Data systems that link student records over time and that link students with their teachers could facilitate value-added modeling and other new approaches to data analysis. Value-added models, in particular, hold promise for improving users' ability to understand achievement trends and evaluate the effects of programs or interventions on individual students, although additional work is needed to evaluate the validity of these approaches (McCaffrey et al., 2003). The development of additional measures of student achievement (e.g., interim assessments administered as part of the regular classroom instruction throughout the year), as well as measures of teachers' instructional practices, could lead to a richer understanding of students' learning growth. Researchers have been active in developing such measures but have not always worked directly with practitioners or other users to ensure that the measures meet users' needs. Foundations could facilitate such interactions by funding projects that link researchers, practitioners, and other users of information about students and schools in collaborative efforts to develop new data systems or measures.

Work with researchers to enhance the usability of results.

Much of the advice for researchers presented above is intended to make research results more useful and accessible to educators and other users of research results. Although many researchers have good intentions with regard to improving the reporting and dissemination of results, it is common for researchers to lack the time, resources, and expertise to act effec-

tively on those intentions. Program officers could help by providing funding for activities that would improve research dissemination. Examples of potentially promising activities include researcher training sessions on effective reporting strategies; funding for collaborations among researchers, practitioners, and the media; and research syntheses that summarize what is known about a topic and present the information in an accessible format.

Conclusions

This paper builds on earlier work (Stecher, Hamilton, & Gonzalez, 2003) to present guidance for improving the implementation of test-based accountability. We have synthesized input from a variety of stakeholders to provide advice that will enhance the utility of information produced by states' accountability systems enacted under NCLB. As states, districts, and schools implement the law, continuing feedback should be sought from all stakeholders—educators, parents, journalists, program officers, and researchers. This feedback can inform changes to the system and facilitate collaboration among these groups in ways that will further enhance the utility of accountability data for decisionmaking. In particular, two external audiences—journalists and foundation program officers—have important roles to play in the success of NCLB and school improvement efforts. Educators and researchers can do more to help them obtain the information they need to contribute to the school improvement efforts set in motion by NCLB.

Bibliography

Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states.* Boston, MA: National Board on Educational Testing and Public Policy.

Colvin, R. L. (2002/2003). Journalists tend to mistrust education research. *Hechinger News,* Winter/Spring. Available at http://www.tc.edu/hechinger/newsletters/Hechinger_News_7-FINAL.pdf (last accessed 11/18/03).

Gootman, E. (2003, September 12). Define paradox? A leading school, below standard. *New York Times.*

Greene, J. P., & Forster, G. (2003). *Public high school graduation and college readiness rates in the United States.* New York: Manhattan Institute for Policy Research, Education Working Paper No. 3.

Hall, D., Wiener, R., & Carey, K. (2003). *What new "AYP" information tells us about schools, states, and public education.* Washington, DC: The Education Trust.

Hambleton, R. K., & Slater, S. C. (1995). *Do policy-makers and educators understand NAEP reports?* (Final Report). Washington, DC: National Center for Education Statistics.

Hamilton, L. S., & Stecher, B. M. (2002). Improving test-based accountability. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (eds.), *Making sense of test-based accountability in education.* Santa Monica, CA: RAND Corporation, MR-1554-EDU, pp. 121–144.

Hamilton, L. S., Stecher, B. M., & Klein, S. P. (eds.). (2002). *Making sense of test-based accountability in education.* Santa Monica, CA: RAND Corporation, MR-1554-EDU.

Jones, G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan,* 81(3), 199–203.

Koretz, D. (2003). Using multiple measures to address perverse incentives and score inflation. *Educational Measurement: Issues and Practice,* 22(2), 18–26.

Koretz, D., & Diebert, E. (1996). Setting standards and interpreting achievement: A cautionary tale from the National Assessment of Educational Progress. *Educational Assessment,* 3, 53–81.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: RAND Corporation, MG-158-EDU.

No Child Left Behind Act of 2001. Public Law 107-110 (January 8, 2002). Reauthorizations to the Elementary and Secondary Act of 1965, U.S. Code Title 20.

Romberg, T. A., Zarinia, E. A., & Williams, S. R. (1989). *The influence of mandated testing on mathematics instruction: Grade 8 teachers' perceptions.* Madison, WI: National Center for Research in Mathematical Science Education, University of Wisconsin–Madison.

Shepard, L. A., & Dougherty, K. C. (1991). *Effects of high-stakes testing on instruction.* Paper presented at the Annual Meeting of the American Educational Research Association and National Council on Measurement in Education, Chicago.

Smith, M. L., Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1991). *The role of testing in elementary schools.* Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, CSE Technical Report 321.

Stecher, B. M., & Barron, S. I. (1999). *Quadrennial milepost accountability testing in Kentucky.* Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, CSE Technical Report 505.

Stecher, B. M., Barron, S. I., Chun, T., & Ross, K. (2000). *The effects of the Washington state education reform on schools and classrooms.* Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, CSE Technical Report 525.

Stecher, B. M., Hamilton, L. S., & Gonzalez, G. (2003). *Working smarter to leave no child behind: Practical insights for school leaders.* Santa Monica, CA: RAND Corporation, WP-138.

Swanson, C. B. (2003). *Keeping count and losing count: Calculating graduation rates for all students under NCLB accountability.* Washington, DC: The Urban Institute.