

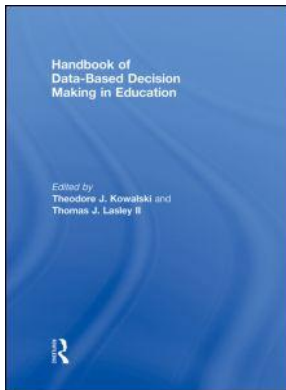
This article was downloaded by: 10.3.97.143

On: 08 Dec 2023

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Handbook of Data-Based Decision Making in Education

Theodore J. Kowalski, Thomas J. Lasley

The Role of Private Firms in Data-Based Decision Making

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9780203888803.ch4>

Patricia Burch, Tracy Hayes

Published online on: 13 Oct 2008

How to cite :- Patricia Burch, Tracy Hayes. 13 Oct 2008, *The Role of Private Firms in Data-Based Decision Making from: Handbook of Data-Based Decision Making in Education* Routledge

Accessed on: 08 Dec 2023

<https://www.routledgehandbooks.com/doi/10.4324/9780203888803.ch4>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Handbook of Data-Based Decision Making in Education

Edited by

**Theodore J. Kowalski and
Thomas J. Lasley II**

First published 2009
by Routledge
270 Madison Ave, New York, NY 10016

Simultaneously published in the UK
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

This edition published in the Taylor & Francis e-Library, 2008.

“To purchase your own copy of this or any of Taylor & Francis or Routledge’s collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk.”

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2009 Taylor & Francis

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Handbook of data-based decision making in education / Theodore J. Kowalski & Thomas J. Lasley II, editors.
p. cm.

Includes bibliographic references and index.

1. School management and organization—Decision making—Handbooks, manuals, etc. I. Kowalski, Theodor 1943— II. Lasley II, Thomas J. 1947—
LB2805 .H2862 2008
371.2 22

ISBN 0-203-88880-4 Master e-book ISBN

ISBN10: 0-415-96503-9 (hbk)

ISBN10: 0-415-96504-7 (pbk)

ISBN10: 0-203-88880-4 (ebk)

ISBN13: 978-0-415-96503-3 (hbk)

ISBN13: 978-0-415-96504-0 (pbk)

ISBN13: 978-0-203-88880-3 (ebk)

4

The Role of Private Firms in Data-Based Decision Making

Patricia Burch and Tracy Hayes

University of Wisconsin-Madison

Since the adoption of the No Child Left Behind Act (NCLB), the 2001 reauthorization of the Elementary and Secondary Education Act (ESEA), the work of schools has been focused increasingly on the collection of data and the use of the data in informing policy decisions (U.S. Department of Education, 2001). Although significant numbers of large, urban districts are purchasing testing goods and services (as suggested by the survey data reported in a subsequent section of this chapter) there is virtually no scholarly research on their use or impact. There are case studies provided by vendors themselves, which are available on their Web sites, but most of these do not appear to be based on systematic, rigorous research. There is a nascent and insightful scholarship conducted by a handful of academics seeking to understand *how* these systems are being used (Halverson, Pritchett, Grigg, & Thomas, 2005; Mandinach, Rivas, Light, & Heinze, 2006). However, these studies are isolated examples.

This chapter aims to contribute to research on the implementation and effects of the current press for data-based decision making through analysis of both the role of private firms as suppliers of the goods and services being purchased to meet the mandates of NCLB and the recent trend toward the adoption of testing technology. We present empirical evidence of widespread district purchase of testing technology since the adoption of NCLB, the cost of services, and patterns in usage. The emphasis of NCLB on test score data has triggered some interesting and potentially useful innovations in the testing market, in particular improvements and needed overhaul in the design of tests themselves. However, on balance, the kinds of goods and services being designed by firms in response to adequate yearly progress (AYP) mandates represent a very partial and potentially costly solution to a deeper set of problems linked to long-standing structural inequalities and institutionalized district practices.

Research Methods

To begin to map the policy landscape of testing technology from the perspective of public agencies and customers, we collected survey data from 28 of the 30 largest school districts in the United States. We drew our sample from the 2003–2004 survey

of the 100 largest school districts as conducted by the National Center for Education Statistics (NCES, 2006). These were the 30 largest districts in the country as measured by student enrollment in the 2003–2004 school years.

The survey data provided broad patterns in district usage, but provided little in the way of contextualized knowledge of how local political, economic, and social histories interact with and shape the implementation of testing technology. To investigate more micro-level contextualized influences, we selected three communities, each of which contracted with a firm that provided assessment solutions, and we conducted case studies in these communities. These case studies included interviews with district administrators, school staff, and firm employees. Based on this combined analysis, we conclude with a discussion of potential benefits and risks in the trends described and offer several suggestions for further developing depth of understanding around these issues and informing action.

NCLB placed heavy emphasis on expanding the scale of testing in K-12 schools, both through mandates regarding the frequency of testing (once a year), the scope of testing (students in all public schools, including English Language Learners and Special Education students), and even the format of tests. With regard to the latter, it required that states design and administer tests in English to students attending public schools in the United States for three years or more. NCLB also introduced a highly detailed set of regulations about how the data should be analyzed, reported, and disseminated (by subgroup, within a certain time frame, and via new reporting formats, to parents).

Benchmark Assessment Systems as a Strategy for Complying with NCLB

Although NCLB provided detailed guidance to local and state governments around testing, devising ways to pay for the tests and their administration remained the responsibility of local and state education agencies. In response to the mandates of NCLB, communities have purchased benchmark assessment systems. We use the term benchmark assessment systems to describe the services and products that districts and states employ to gauge student performance on standardized tests system-wide prior to the annual state level summative exams that are used to measure AYP under NCLB. The cost structure of benchmark assessment systems is different from the textbook and testing purchases of the past. Rather than actually purchasing a device as in the case of a textbook, typically districts lease the system for one to three years. The basic contract is for the database shell—the application that allows the district to store test results and generate and disaggregate data. The district will sign up for a software license and an initial year of support and services to implement the system. Firms typically charge about 20% of the license for the support services which include telephone support and updates for the product. These costs, as described later in the chapter, appear to range from \$5 to \$15 annually per student. In addition to the basic contract, districts may decide to purchase additional supplementary components such as professional development, kits for customizing analysis of data in ways that are aligned with state or local policies, and software for generating different kinds of reports.

As current subscribers, districts may receive free software updates from the contractor as they become available. However, these updates tend to be automatic, and every district with a contract with the firm receives them. New software typically is loaded onto the district server whether the district desires it or not. When districts lease the software, issues of ownership over data can also become more blurred. In the past, assessment data (in paper form) would be housed in the district offices—typically under the jurisdiction and lock and key of offices with titles such as Research and Evaluation. In the current structure, data can be housed on a server managed and operated by external providers. Firms that sell assessment software offer these services at an additional charge. The security of these data is the responsibility of the firm, although the confidentiality of the data is governed in part by local and federal policy.

The cost structure of the system has implications for public access and ownership of the data. When districts purchase textbooks, they may not have control over the design of the content; yet through the contract, they acquire a physical asset—the textbook itself, which they keep even if they decide to adopt a different textbook series from a different publisher. In contrast, when they contract with private firms of testing software, districts do not own the software that is used for managing the requirements of NCLB and its instructional goals. They borrow it. They essentially lease the shell or software program that allows them to manipulate the test score data. If they terminate the contract, they no longer have access to the shell and may not have access to the data.

Preliminary Trends in District Purchases and Usage

To begin to investigate local perspectives and responses to industry developments, district officials in 30 large urban school systems were surveyed. Survey questions were designed to map the extent to which large, urban school districts are purchasing benchmark assessment systems in order to meet the mandates of NCLB and to begin to identify where the pressure for these purchases originates. The sample was limited to all 30 school districts (excluding Puerto Rico and Hawaii, which are state-run organizations) with a student population greater than 94,000 according to the NCES report, *Characteristics of the Largest 100 Public Elementary and Secondary School Districts in the United States, 2003–2004*. As reflected in Table 4.1, we received a response from at least one official in 93.33% of the 30 school districts surveyed. Out of a total of 148 people across those districts, 54 (36.54%) responded. The people most likely to respond were those in the role groups of Technology (32.69%) and Research, Assessment, and Accountability (32.69%), although we also received responses from those in Title I Administration (15.38%), Budget and Finance (13.46%), and Curriculum and Instruction (9.25%). There are several primary patterns suggested by the survey data.

Table 4.1 District survey: Characteristics of sample.

Response rate (districts)	93.33%
Response rate (individual respondents)	36.54%
Percent of responding districts by region	
South/Southeast	53.57%
Northeast	21.42%
West	14.29%
Midwest	10.71%
Percent of individual respondents by role group	
Technology	32.69%
Research, Assessment, and Accountability	32.69%
Title I/Supplemental Educational Services	15.38%
Budget and Finance	13.46%
Curriculum and Instruction	9.25%

Widespread Use of Benchmark Assessment Systems

Large urban school districts are purchasing testing technology and doing so in part to meet the mandates of NCLB. Over 82% of the responding districts indicated that their district currently invested in the testing technology commonly known as a benchmark assessment system that was intended to allow them to administer periodic assessments and score, disaggregate and report the results to stakeholders in the district. Sixty-nine percent (69.57%) of these districts stated that they had purchased their system since the passage of NCLB. Five districts (21.74%) purchased their systems before NCLB. Two districts (8.7%) were unaware of when the system was purchased.

While NCLB is an important factor, district officials reported that the pressure to purchase these systems is coming from multiple directions. This includes pressure from school boards and superintendents (73.91%) and state mandates (52.17%). A much smaller percentage identified the pressure to purchase the system as emanating from pressures *within* lower levels of the system including other district staff (13.04%), school administrators (8.69%), and teachers (21.74%).

Improving instruction across grade levels appeared to be a prominent objective for purchasing the system; simplifying accountability reporting was not. By and large, the two most frequently mentioned goals were improving instruction (47.83%) and predicting student progress (17.39%). Employees from several of the districts gave differing responses, but all of those responses were the same as the two mentioned above. No district or individual respondent selected simplifying accountability reporting as a primary goal for the system.

These systems are typically implemented across all grade levels at most schools within each district. Eighty-two percent of districts which have a system report that it is in place at all levels: elementary, middle, and high school. Only one district indicated that the system had been installed in only the elementary and middle grades. In addition, nearly two-thirds of districts had installed the system at 76 to 100% of their schools.

Officials from five districts gave answers which varied from their co-workers when indicating what percentage of their schools had installed the system. This may indicate an ongoing, phased rollout of the system. Indeed, when asked to choose an adjective to describe the speed of system implementation, employees from 26.09% of districts all selected “phased.” The discrepancy in answers between some colleagues also points to minimal involvement in implementation by employees of certain departments. This conclusion is further supported by the fact that a discrepancy in answers also occurred between officials in 39.10% of our districts when they were asked about implementation speed.

Many Suppliers; No Apparent Monopoly Yet

Interestingly, although many of the large vendors of this technology claim to be working with hundreds of school districts (Burch, in press), they were a very small presence among this group of large, urban districts. The most frequently mentioned products were Edusoft, a product of Riverside Publishing (a Houghton Mifflin subsidiary), and The Princeton Review, which is a public company specializing in standardized test preparation and K-12 solutions. However, Edusoft is currently in place in only three out of the 23 districts with systems, and Princeton Review was mentioned by only four districts. Additionally, 50 to 75% of those districts who mentioned either Edusoft or Princeton Review also indicated other systems. About 10.53% of officials indicated that the district’s prior relationship with the provider was the main reason for their district’s choice of provider. Therefore, although familiarity with a provider appeared to have no effect on a district’s decision to buy a benchmark assessment system in the first place, it did give some of the major players a slight advantage over others once the district had made the choice to buy.

Cost of Systems

District officials across role groups appeared to have limited information about the costs of the systems. However, as reflected in Figure 4.1, of the 15 districts that were able to answer this question, all of them paid under \$15 per student per year, and the majority (80%) paid under \$10 per student per year. If these estimates were to apply to all the districts in the sample, the base cost of a system (without add-ons) per student per year would vary within the approximate ranges of \$200,000 to \$500,000 for the smallest district, and \$1.4 million to \$4 million for the largest.

The full costs of these services and products may be much higher given that 24 of the 28 districts reported purchasing additional components from vendors and that in interviews executives described setting up software licenses in ways that required customers to make additional purchases over time as part of the contract. Over 43% of districts purchased services marketed as helping them customize assessments, while 30.43% of districts reported that they purchased additional question item banks. Over 20% of districts bought professional development and over 25% bought a service contract. Only four districts did not buy any additional components, either

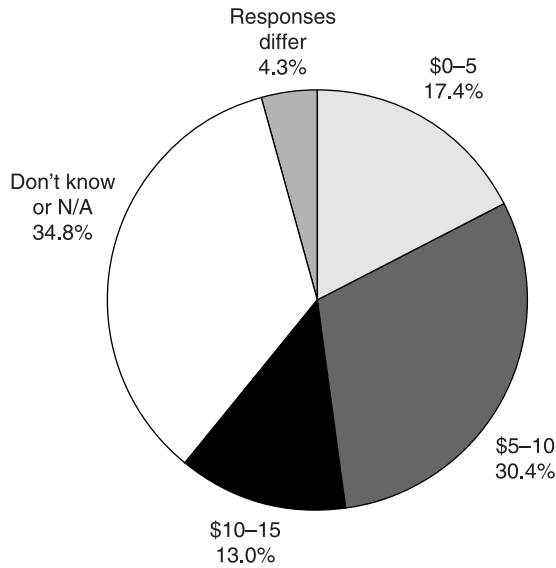


Figure 4.1 Cost of assessment systems.

Note: Based on responses from 23 districts.

because they were still investigating options or because they built their own systems and therefore did not have to “purchase.”

Relative to district spending as a whole, the amount of money being spent on these systems is small. In our sample, the range of per-pupil spending falls between \$6,000 and \$11,000 per student per year. Thus, given the cost per student estimated above, benchmark systems represent approximately 1% of what districts spend on students every year. The more fundamental issue, which requires additional research, concerns the indirect and hidden costs to districts for these purchases, as suggested from our case study data. This includes the significant costs involved in the development and administration of the Request for Proposal process, and the training and capacity building of firms that may have little experience in school districts or in the education market generally.

There appears to be a lack of connection between the policy impetus for buying the system and the funds used for the purchase. Even though external forces (policy pressures emanating from outside of the district) are identified by districts as driving purchases, districts are relying heavily on district general funds in order to cover the cost of the systems. In the survey, 60.87% reported drawing on district general funds. However, only six (26.08%) districts listed funds from ESEA as having helped to pay for their system, and a mere 13.04% drew on state funds (see Figure 4.2).

Varying Perceptions of Usage

In the survey, we also asked district officials for their assessment of the percentage of teachers, school administrators, and district officials who they believed were active users of the benchmark assessment systems (see Figures 4.3 and 4.4). Perceptions of

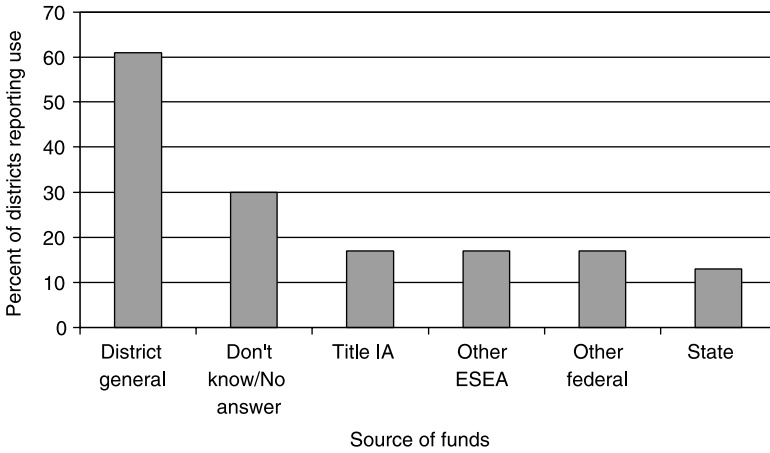


Figure 4.2 Funding streams for purchase and implementation of assessment systems.

Note: Based on responses from 23 districts. Respondents could choose multiple sources of funding.

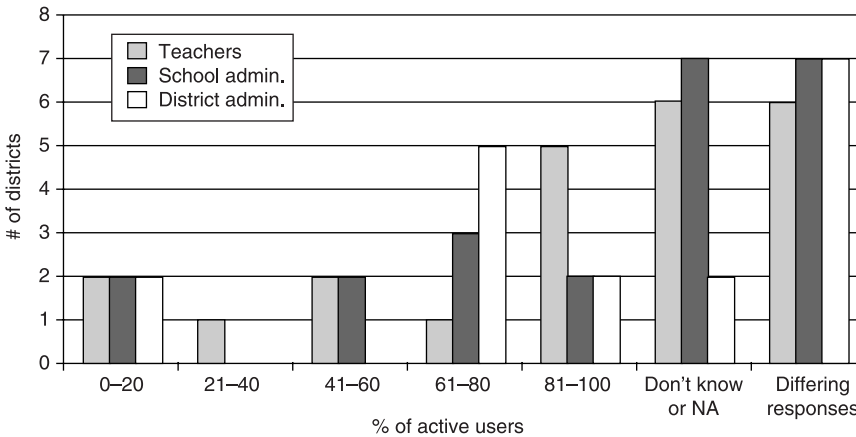


Figure 4.3 District perceptions of active use of the benchmark assessment system among teachers, school administrators, and district administrators.

Note: Based on responses from 23 districts.

active use differed dramatically both between and within role groups. The majority of respondents felt that less than 80% of district administrators were active users. As a group, respondents were much more likely to identify higher numbers of active users among both school administrators and teachers. A full 27% of respondents had no idea how many employees in any role group were active users of the system. When disaggregated by district instead of by official, the responses also show that perceptions differ not only between districts, but within them. Within very few districts were respondents in agreement over what percentage of teachers, school administrators, and district administrators were active users of the system. All respondents within 30.43% of districts had no idea how many school and district administrators were active users, and respondents within another 30.43% of districts disagreed with one another as to the percentages. There was only slightly less uncertainty when it

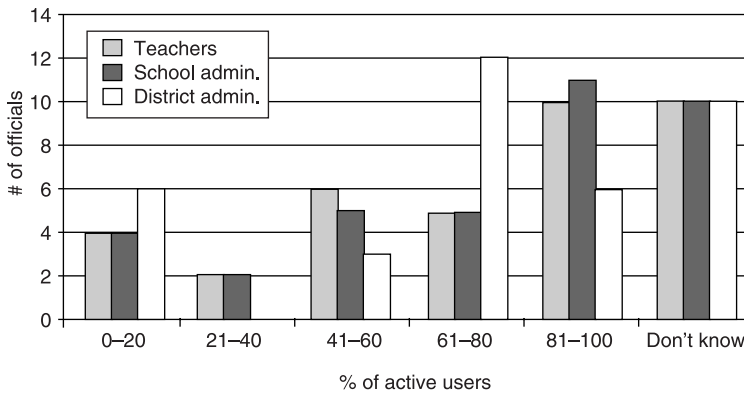


Figure 4.4 Officials' perceptions of the percentage of "active users" among teachers, school administrators, and district administrators.

Note: Based on responses from 37 officials across 23 districts.

came to perceptions of teacher usage. The primary conclusion that the data on usage support is that there is often not one prevailing perception of teacher, school administrator or district usage within or across districts.

These preliminary patterns raise several potential trends around the purchase of benchmark assessment systems in the move toward data-based decision making. We briefly identify these trends before exploring them in greater depth in the case studies that follow. First, the survey provides further evidence of the role of NCLB as a supporting factor in the selling of K-12 assessment services and goods. The data from the survey suggest that the pressure to buy is coming from multiple levels of the system, not just federal mandates. The pressures include superintendent/school board (72%), state mandates (50%) and NCLB (40%). These patterns appear to contradict or at least add nuance to assertions that districts' decisions to contract emanate from or are based primarily on internal needs or capacity. Here, as in other aspects of the education market, the pressure to contract is inextricably linked to the external policy environment (Burch, 2006).

Second, the survey data suggest evidence of loose coupling between district practice and private firms' policy intents. When applied to education, loose coupling generally refers to the gap between policy goals and administrative practices. Here, we are suggesting that loose coupling can also occur between private firms representing policy goals and the interests and practices of public agencies. While the marketing materials of firms emphasize economic principles of cost cutting, time efficiency, and ease of reporting, district officials intend the systems for a different set of purposes. They aim to use them to improve instruction and refer infrequently to the firms' stated objectives of streamlining reporting.

Third, those officials responsible for overseeing the implementation of these systems appear to possess either conflicting or limited knowledge about critical aspects of the systems. The systems are designed for use across multiple district departments and yet officials within districts appear to have limited and contradictory information about usage and cost. This trend is important because districts may actually need more support than they are getting in implementing the systems. In addition, if district officials

cannot agree on usage figures, how can they judge whether the implementation has been successful or whether the system has managed to help them improve instruction?

Benchmark Assessment Systems in Local Context

We next consider the locally situated meanings and significance of the trends through three case studies of district practice. The first district, which we call Beachside, was a very large, decentralized urban school district with over 200,000 students. The second district, which we call Carbondale, was a much smaller urban school district with a student enrollment in the range of 40,000 to 50,000 students. The third district, which we call Chesterville, also was urban. It fell between Beachside and Carbondale in terms of student enrollment, with just over 100,000 students. During the school year in which data were collected, Carbondale was not yet in program improvement. In contrast, Beachside and Chesterville have both been in need of improvement since the 2004 to 2005 school years.

In each district, we conducted extended interviews with several district officials across different departments from January 2007 to August 2007. Data collection in each site included repeated semi-structured interviews with district officials and extensive document review. In one district (Carbondale) we supplemented interviews and document review with shadowing of district officials and observations of school level usage of benchmark assessment systems. In our primary case (Carbondale), we interviewed three staff members in the Department of Curriculum and Instruction, two staff in the Office of Technology and Instruction, and the Director of Assessment and Accountability. In Beachside, we interviewed the Director of Technology Instruction, the Director of Evaluation and Research and the Director of Curriculum and Instruction. In Chesterville, we spoke with the current Chief Academic Officer and Executive Director of Technology Instruction, and the former Director of Middle School Programs.

Each case focuses on the district level functions and challenges emerging in districts under pressure to make test score targets under NCLB. We start with a vignette of the implementation of Test Whiz in our primary case Carbondale and then compare this, through a narrative of broader brush strokes, to the implementation of Test Whiz in Beachside and Chesterville. All three case studies illustrate how power differentials within district organizations are activated by the purchase of these systems and the complex issues around equity and access that these “solutions” engender. Here, as in other forms of top-down policy, context still matters. Whether these systems are used and how they are adapted is contingent on local situations including the strength of unions and the size of the district.

Carbondale School District

The school district of the city of Carbondale enrolls approximately 40,000 to 50,000 students and is located in the western United States. Approximately 6,000 to 7,000 of the students in the district are ELL students and 4,000 have Individualized Education

Plans. Close to half the students enrolled in Carbondale identify as Hispanic or Latino. Approximately 30% of students enrolled are African American, and 10% are Asian. Over half of the residents in the city of Carbondale itself are White, 40% are Hispanic and Latino and 8% are African American. Approximately 16% of Carbondale's population lives below the poverty line. The school district was placed under court ordered desegregation in the mid-1960s. Most of the city's poverty and crime are concentrated in one area of the city—which is also largely African American and Latino. This neighborhood also has the highest unemployment rate in the district and lowest educational rate as measured in terms of number of adults over 18 with high school diplomas.

Sharon K has been a K-12 Mathematics Instructional Specialist in Carbondale since 1998. In her words, this meant that she was responsible “for all things mathematics” in the district, from overseeing textbook adoption every seven years, to organizing staff development for teachers across grade levels, to analyzing math scores for the district. Prior to her current position, she worked as a product development manager and before that as an elementary school principal. She was one of three African Americans working in the district at the level of professional—Master's degree and higher. She was raised “poor and Black” in central city Los Angeles and moved to Carbondale to attend college and obtain her teaching certification. She has lived there ever since.

Access to Data In 2002, the district purchased a benchmark assessment system called Test Whiz. According to Sharon and others interviewed in the district, Test Whiz offered the district many advantages in terms of facilitating district management of large amounts of data. Before Test Whiz, teachers would score common assessments by hand. Staff would then get together at each school and with pencil and paper attempt to identify problem areas. There was little capacity for item analysis based on skill strands and little capacity for the district to develop a picture of student performance districtwide. Sharon found Test Whiz to be a useful tool for analyzing the added value of her own work and that of her staff and making more informed decisions about where to invest resources. On occasion, she would analyze the test scores of students in schools and classrooms in which her staff had done intensive staff development. In other instances, Sharon used Test Whiz to analyze the efficacy and equity of particular items on district level assessments. She would then use these data to assess students' difficulty with those items on a districtwide basis in relation to other items, as another gauge for where to focus staff development.

Further, Sharon and others in the district reported that Test Whiz had helped decentralize access to data within the district. Specifically, it made data more available to staff in Curriculum and Instruction, rather than being the exclusive domain of staff in Research and Accountability. Before the district began using Test Whiz, if Sharon was interested in comparing the performance of certain schools to others in the district (for example, schools with higher proportions of White students to schools with higher proportions of ELL and African American students) she would have to go through the Office of Research and Accountability—the four-person department that occupied a small room on the second floor of the district offices. It

could be weeks or even months before Sharon could expect an answer. With Test Whiz, there was less paper and Sharon could simply conduct the analyses herself.

Raising Attention to Equity Issues at District Level Perhaps the thing that Sharon liked most about Test Whiz had less to do with Test Whiz and more to do with NCLB. While the district had a long history of collecting data—and like most districts, had an entire department devoted to this purpose—it rarely did anything with the data, according to Sharon. The data might show that there were significant differences in the academic performance of subgroups in the district by race or socioeconomic status, but from Sharon’s perspective, the district rarely felt compelled to do anything about the gaps. The data were available (although time-consuming to obtain) but even after being analyzed “they just sat there.” Sharon referred to her district’s approach to data before NCLB and Test Whiz as “living on the average.” It paid attention to the kids in the middle and ignored the sharp differentials between the scores of children at the high and low ends of the scale. In a district which one of the few African American district employees characterized as largely indifferent to the needs of students of color, Test Whiz made available to the district for the first time periodic (rather than annual) data of student performance by race and income level.

In other ways Test Whiz seemed to Sharon and others to be disconnected from the core issues of equity and access in the district. One core issue involved teachers’ low expectations of students of color and students living in poverty. Throughout her years in the district, Sharon had become increasingly convinced that ELL and African American students in the district were not being given the same opportunities to learn challenging content in spite of state standards and in spite of the fact that the district was no longer under a court desegregation order. Specifically, early in her career, while teaching at a low-income, high-ELL, high-minority school, she had been at first surprised at the amount of improvement her students would make and then eventually realized after consulting with the students’ teachers from past years that they had not been held to grade level expectations before, but had been given a watered down curriculum. Later, as a curriculum staff developer, she recalled walking through the campus of a school that had consistently low test scores for ELL and African American students and while visiting the classrooms, “seeing that not a single classroom in the school was teaching on grade level and that students of color and ELL students were assigned to what was quietly referred to as ‘non-regular classes.’ ”

For individuals like Sharon, who know the district and have committed their lives to it, the achievement gap stems primarily from lack of parity in students’ opportunities to learn rather than inefficient district organization or lack of good data. This point hit home for Sharon last year while working with the group of schools that had not made proficiency under NCLB and were therefore deemed in need of improvement. In the course of Sharon’s work with the schools it became clear to her that teachers in the second and third grades were not teaching fractions in ways defined by grade level standards. Analysis of test scores (powered by Test Whiz) had revealed low levels of proficiency in math for second- and third-grade students in the school, particularly as compared to the performance of second and third graders in a neighboring school that had a much smaller proportion of students from a low

socio-economic bracket. The teachers interpreted the results to mean that the material was too hard for the students. When Sharon discussed the issue with the teachers, they told her they felt that their students (the majority of whom were Black, Latino and very poor) could not handle the content defined by grade level standards. In response, they had adjusted the curricula and adopted their own vision of what it meant to teach fractions at grade level standards for their students.

Thus, tools such as Test Whiz can assist in the goal of more democratic schooling by helping districts disaggregate data by subgroup. However, particularly in districts such as Carbondale, with long legacies of institutional racism, benchmark assessment systems appear to have little effect on deeply entrenched race- and class-based biases about who can achieve. In a policy climate where the rhetoric is heavy on “leaving no child behind,” benchmark assessment systems hold out the promise of helping all children have access to challenging content and to achieve. However, as in the case of any other externally designed “reform,” benchmark assessment systems are implemented into existing routines and systems of practice. This includes both systems of practice that may *enable* deep change in instruction as well as routines and deeply entrenched stereotypes that impede equal access and opportunity to challenging curriculum.

Beachside School District

Beachside is a much larger school district than Carbondale (over 200,000 students as opposed to 45,000 students). The City of Beachside is a large, densely populated area. It is a very diverse city in which 49% of residents are White, 10% are Black, 11% are Asian, 49% identify as Hispanic, and 27% identify as other ethnicities. Beachside School District serves one of the largest and most diverse populations of students in the country. The characteristics of this student population stand in marked contrast to those of the residents: 73% of Beachside students are Hispanic (not White or Black), 11% are Black, only 9% are White, and just over 6% are Asian, Pacific Islander, or Filipino. Almost 82% of Beachside schools are Title I eligible, and 64.4% of Beachside students participate in schoolwide Title I programs.

Although it has consistently increased the percentage of proficiency among students in most subgroups over the past four years, the district has still struggled to meet the mandates of NCLB, and in fact has never made AYP itself. Additionally, while some subgroups have met the criteria for AYP year after year, no subgroup is anywhere near 100% proficient in either language arts or mathematics. It is a highly decentralized district, separated into regions, with subject area coordinators designated for each region.

Beachside started experimenting with benchmark assessments in 2000, when the district adopted a curriculum for elementary reading that was offered through the state and included periodic assessments. This curriculum represented the first time that the district had adopted standardized tests that were to be used across all elementary schools. As one Beachside official notes, “prior to that, mostly, schools were using teacher-developed materials, or test prep, or end-of-chapter tests, but nothing was done in a systematic way to collect data.” The assessments were given and then

hand-scored by teachers. To enable the data from the assessments to really be of use to teachers and principals, a district employee developed a basic spreadsheet program into which the test data could be inputted. It eventually developed into a system that was connected to the district's student information system and would indicate in reports the teachers could print out which students were progressing as expected, which students were starting to struggle, and which were at a critical point of need.

In mid-2001, a new superintendent arrived at the district with a very strong view of where it should be headed. His theory of action was geared toward implementing a standards-based curriculum, improving instruction through professional development, school-site coaching for teachers, strong positive leadership, and periodic assessments. It was therefore time to expand the benchmark assessment system to include more grade levels and subject areas. Unfortunately, the district's developer, who by that time had gone out on his own, did not have the capacity to do it. The district had also been warned in an audit against building too many of its own structures from scratch because it could cost more time and money than buying products from outside vendors. So instead, they contracted with Test Whiz to design and write a series of benchmark assessments for multiple grade levels. The following year, Beachside later re-contracted with Test Whiz for three years of additional service at a fixed cost, with an annual option to renew.

The terms of the contract were discussed and agreed to by both sides beforehand. However, according to district officials, neither Test Whiz nor the vendor that succeeded Test Whiz was prepared for the size and complexity of the district. As reported by one district official, "The system was unusable for more than a year. They [vendor] were not prepared for the complexity of our district." Beachside district staff had to spend an extraordinary amount of time teaching Test Whiz about the district and helping the company adapt its products to the district's unique needs; for example, the fact that the district had year-round schools. After three years, the district signed a new contract with a new vendor. According to several administrators, the district then had to spend a whole year backtracking. As one individual commented, "We had to start all over with the new company."

In its bid for a renewed contract, Test Whiz had banked on the district paying a much higher price. The new price estimate was 50% greater than what the district had been paying (\$6,000,000 versus \$9,000,000). In addition, while Test Whiz had grown to know Beachside well, there were still aspects of the system that were not working. Specifically, under Test Whiz, teachers were responsible for scanning the test score data themselves, but the equipment that the district had purchased was not up to the demands of the system, and Beachside's strong teachers' union strenuously objected to this use of teachers' time. Furthermore, as in the case of Carbondale, the individuals we spoke with viewed the system as deeply disconnected from the deeper set of changes required if the achievement gap was to be addressed in any significant way. The Chief Instructional Officer in Beachside put it this way:

Changing instruction is the hard work. Putting periodic assessments into the system is a snap compared to the larger issue of helping people understand how to use this *[sic]* data and how to create the conditions for teachers to help teachers be reflective about their practice. That's professional development in itself. It's conversations about removing barriers, it's about the

principal being present and observing their *[sic]* teachers. It is examining the quality of a grade level meeting and pressing on that quality. It's not letting people off the hook simply because they gave the test but getting them to look at the data. No, that is the hard work before us.

As in the case of Carbondale, Beachside officials also referenced the role of racist and class-based stereotypes, and a district culture “where it has been OK to look at data through the lens that it is the kids have the problem” rather than looking critically at district and school level policies and practices.

The implementation of the formative assessment system also depends on the strength of the union in the district and its involvement in the process. In Beachside, union officials appeared to have considerable input into the planning and use of formative assessment systems; specifically, when the assessments would be given, and whether teachers or the vendor would be responsible for scanning answer sheets. The union was adamantly opposed to an approach where teachers scanned the test scores themselves. Although the district had little sense beforehand that there might be an uproar, the strength of the union is likely one reason why Beachside was eventually forced to establish a contract with a vendor that provided scanning services rather than relying on teachers for this work. Part of what won the second vendor the contract was its solution to this issue, which involved having FedEx pick up answer sheets from each school and return them to a central location for scanning and uploading to the system.

The Beachside case further reveals the complex web of district policy and practice that intersect with the technological tools being introduced by private firms in the context of NCLB. As in the case of Carbondale, district officials in Beachside viewed contracts with firms like Test Whiz as one slice of what necessarily must be a larger set of system changes that involved professional development, collaboration and (over time, in part because of pressure from the union) respect for teachers' primary responsibility for classroom instruction. In Beachside as in Carbondale, effective implementation of the system required substantive learning and change on the part of the firm itself. The ability of firms to learn and adapt to diverse district contexts helps frame and define district level usage of the system and ultimately, its interactions with teachers around the system. At the same time, district context must play a large part in determining whether to lease a product off-the-shelf from an established firm, or to enter into partnership with a developer to build a system from scratch. This is illustrated by the decision made by our third case, Chesterville School District, which decided to create a system in-house tailored to its specific needs.

Chesterville School District

The Chesterville School District is a large, urban district with over 100,000 students. Along with the community that surrounds it, Chesterville has undergone sustained growth in recent years; doubling in student population since 1990. Unlike in Carbondale and Beachside, the majority of the students in the district are White. The percentage of Chesterville students enrolled in Title I programs also stands in marked contrast to both Carbondale and Beachside, although it has jumped dramatically with the increase in population. Much like Beachside, Chesterville has never made

AYP. Indeed, the percentage of schools making AYP within Chesterville has dropped steadily and dramatically over the past four years, from a high of nearly 75% in school year 2003 to 2004 to the current low of roughly 40% in year 2006 to 2007.

For years prior to 2001, Chesterville was what some researchers call “data rich and information poor” (Dufour, Eaker, & Dufour, 2005; Halverson et al., 2005). For example, in the elementary schools, teachers created detailed, handwritten profile cards for each student, on which they collected information about that child’s performance in mathematics and literacy. However, as one official noted, “other than spreading all these cards out on the living room floor, there was really no way of grouping children to make informed instructional decisions.” Each level of schooling was also at a different point with regard to the collection of data. While the elementary schools had their profile cards, the middle schools were referred to by some as “assessment-free zones” in which no formative data whatsoever were collected.

An initiative called Project SUCCESS was introduced to low-performing elementary and middle schools within the district beginning in school year 2001 to 2002. Project SUCCESS bills itself as a comprehensive school improvement program with seven components, one of which is “evaluation and outcomes.” Chesterville implemented new curriculum and periodic assessments as part of the initiative. Project SUCCESS did not come with a benchmark assessment system, and as a result, the district hired a developer to produce a basic program to track the scores from the tests. The tests were all taken by hand and had to be run through a Scantron machine to be scored, and then the data had to be transferred to the database program and manually manipulated. Thus, there was a very slow turnaround of information, and the technology was in use only by schools that were struggling. By the start of the 2003 school year, only three out of 27 middle schools were participating in Project SUCCESS.

It was at this point that the Director of Middle School Programs decided that she needed a way to track and improve performance in her middle schools across the board, and seized upon the already existing Project SUCCESS assessments as a good way to start. However, she also wanted a more complex database system that would allow the tests to be taken online and an updated, improved capability for analysis and additional reports.

She began to investigate off-the-shelf assessment systems, but what she found was either prohibitively expensive or did not allow for the district’s existing Project SUCCESS assessments to be used. Additionally, she needed something that would allow for the evolution of her program, and as another official noted, “if you were dealing with a commercial product, they would never change it. You’d have to adapt what you did to the commercial product.” So instead, Chesterville officials began a discussion with the programmer who had built the program used to record scores for the Project SUCCESS schools, and asked if he could expand its capabilities and make the assessments available online to all students, and not just to those in struggling schools. Not only was he able to expand it, but the work he did cost much less than any estimates Chesterville had received from off-the-shelf vendors. Although Chesterville still leased the system from the developer, two of the officials we spoke to estimated that the system now known as Gold Star costs them under \$5 per student per year.

Over the past three years, local demand for Gold Star skyrocketed. The system itself has evolved through the suggestions of teachers and central office staff members, who began to request adjustments and additions that would make it easier for them to manipulate and understand the data they collected through the assessments. Staff from all levels of the district also participated in expanding the range of assessments that was available, both by writing items and by participating in the item review process. It is now a comprehensive benchmark assessment system, similar to off-the-shelf products, yet customized to Chesterville's specifications. The system is widely used in elementary and middle schools, and has begun to expand into the high schools.

Although Gold Star is similar to Test Whiz in many ways, Chesterville's experience is distinct from that of Beachside and Carbondale. Officials spoke very positively of the responsiveness of the developer, of Gold Star itself, and of the changes in practice the system had helped to facilitate. For example, although there were initial technical problems with Gold Star, they were fixed very quickly after a simple series of visits the programmer made to schools using the system. Further, in speaking of the effects of the system, one official noted that it had actually helped teachers to break down some of their existing stereotypes. Because of the immediate availability of data, and the ability to see individual item analysis, teachers began to realize that kids who had been previously assumed to be of "low ability" were actually doing much better on some objectives than kids assumed to be of "higher ability." The system helped teachers not only change perceptions, but also understand how to build up support for students in crucial areas.

Chesterville's positive experience with Gold Star might be attributed to the fact that the impetus for the system was somewhat organic and came from the ground up rather than externally. Further, the fact that Gold Star was built and not bought off-the-shelf allowed for input from a much wider variety of school level stakeholders, and may have contributed to less initial resistance toward the system and a quicker adoption.

However, it would be a mistake to assume that building a system is the right choice for every district, or even for the other two in this study. Chesterville's unique structure and context allowed for this option in ways that neither Beachside's nor Carbondale's would have. For instance, with Beachside's strong and active union playing a major role in all teacher-related business it is unlikely that building a system would have been as quick or easy for its officials. The district might have had to formally negotiate the building of the system in a way that Chesterville, which has no teachers' union, did not. Similarly, in Carbondale, a district which had not struggled as much with its AYP goals, the local demand for data might not have been as urgent as it was in Chesterville, a district whose AYP performance had worsened every year since that standard's initiation.

Additional evidence to support the idea that customization is not always best can be found in the cost structure of the systems in Beachside and Chesterville. Both districts paid an annual licensing fee for the use of their systems which was based on the number of students and teachers, and which officials in both districts said fell in the range of \$0 to \$5 per student annually. Beachside's system came with a long list of different report formats that were available to teachers and administrators, which

would reveal different things about the data. Chesterville, on the other hand, had to pay an additional fee for each new report officials wanted. Thus, rather than being available at a fixed price, the annual cost of Chesterville's system varied based on the changes the district made to it each year. In a district the size of Beachside, a fluctuating annual cost may not have been worth the level of customization it offered, especially if it meant sacrificing other benefits of choosing a large, established vendor—such as the unique scoring system Beachside's second provider ultimately offered. Ultimately, the decision of whether to build or buy a system appears to have no obvious or consistent answer, and again must depend on an analysis of each district's structure and available resources.

Benchmark Assessment Systems in the Era of NCLB

The events in these three districts cannot be assumed to generalize to all districts, but they do suggest some problems in the underlying assumptions of NCLB. First, sanctions and mandates may create the incentive for districts to test more frequently, to include special education students in the testing, and to look at data. However, on their own, they do not provide answers on how districts should address these problems, nor do they provide any real indicators of why these problems exist in the district in the first place.

Second, the best-designed, most efficient assessment services and devices are being introduced into social contexts where the people in charge of the system and the people using the system have deeply rooted expectations and stereotypes of what children can achieve. Even when the new technology is installed democratically so that every school has a log-in, even when the tool itself is keyed to local goals in its alignment to standards, even when the technology “works” and allows districts to manage data in a way that they never have before, the overall problem of uneven outcomes and expectations will remain unchanged. In a district like Carbondale where Test Whiz had been implemented across many schools, there were still students left behind, students whose teachers were excused from looking critically at their own practice, on the basis of assumptions about what students placed in special needs classrooms could achieve.

Third, once the decision has been made to use a benchmark assessment system, there is no easy or standard way to determine whether the most cost effective and productive choice would be to construct a system from scratch or lease one off-the-shelf. While constructing a system may win over stakeholders and lead to increased equity and changed perceptions of students as it did in Chesterville, it may also be impossible given the political constraints and the resources of the district in question. A district may spend millions of dollars implementing a system only to realize that there is a seemingly insurmountable roadblock standing in the way of full adoption.

In summary, the stories that unfolded in the three districts suggest the need for considerable caution on the part of districts racing to install benchmark assessment systems in order to meet the mandates of NCLB. For the testing industry, this is a very exciting time as the mandates of NCLB and other policy pressures create incentives for local purchases. The problem is *not* that private firms are involved in the

design and delivery of testing and measurement services. These firms do bring to the table expertise that districts may lack. For example, they can assist in the development of tests that more accurately measure what students know and can do. Yet at the same time, these firms are currently not assisting districts in evaluating the full implications of their decisions, and are not engaged in assessing whether their product is really the right choice for each particular district. If a district is not able to muster the resources or make the changes required of practice at all levels, it runs the risk of allowing the assessment system to become a mechanism for unnecessary summative exams. Finally, as a strategy for helping schools and districts develop a more fine-grained and nuanced picture of their performance, draw appropriate conclusions from the data and use the data to revise district strategy, benchmark assessment systems hold considerable promise. Thoughtful attention by districts of the appropriate role of for-profit firms in developing and managing these systems is important. This should include consideration of the importance of developing internal district capacity, sustaining the work over time, and employing the data in the service of urgent equity issues.

Note

Portions of this chapter also appear in Burch, P. and Hayes, T. (2007) and Hayes, T. (2007). Burch is an Assistant Professor in the Department of Educational Policy Studies at the University of Wisconsin-Madison. At the time of writing, Hayes was a graduate student in the Department of Educational Policy Studies.

References

- Burch, P. (2006). The new educational privatization: Educational contracting and high stakes accountability. *Teachers College Record*, 108(12), 2582–2610.
- Burch, P. (in press). *Hidden markets: The new education privatization*. New York: Routledge.
- Burch, P., & Hayes, T. (2007). *Accountability for sale: Testing, markets and NCLB*. Madison, WI: Center for Education Research.
- Dufour, R., Eaker, R., & Dufour, R. (2005). Recurring themes of professional learning communities and the assumptions they challenge. In R. Dufour, R. Eaker, & R. Dufour (Eds.), *On common ground: The power of professional learning communities* (pp. 7–29). Bloomington, IN: Solution Tree.
- Halverson, R., Prichett, R., Grigg, J., & Thomas, C. (2005). *The new instructional leadership: Creating data-driven instructional systems in schools* (WCER Working Paper No. 2005–10). Madison, WI: Center for Education Research.
- Hayes, T. (2007). *Sea change: The proliferation of benchmark assessment systems in U.S. school districts*. Unpublished Master's thesis, University of Wisconsin, Madison.
- Mandinach, E., Rivas, L., Light, D., & Heinze, C. (2006, April). *The impact of data-driven decision making tools on educational practice: A systems analysis of six school districts*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- National Center for Education Statistics (2006). *Characteristics of the 100 largest public elementary and secondary school districts in the United States: 2003–2004* (NCES Rep. No. 2006–329). Washington, DC: Author. Retrieved December 19, 2006, from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006329>.
- U.S. Department of Education (2001). *NCLB overview*. Retrieved June 15, 2007, from <http://www.ed.gov/nclb/overview/intro/execsumm.html>.