2012

Formative Assessment and Benchmark Testing: Phase 2

Lisa Abrams
Virginia Commonwealth University, lmabrams@vcu.edu

Divya Varier
Virginia Commonwealth Univeristy, varierd@vcu.edu

James H. McMillan
Virginia Commonwealth University, jhmcmill@vcu.ede

Formative Assessment and Benchmark Testing

Phase 2

Prepared by:

Lisa Abrams
Associate Professor
Foundations of Education

Divya Varier
MERC Research Assistant

James McMillan
Professor
Foundations of Education

Table of Contents

List of Tables

Introduction

As school districts respond to test-based accountability requirements the emphasis on using data to drive decision making has most recently focused on using interim or benchmark assessment results. The use of these assessments to monitor student progress and inform instruction with the aim to improve learning is widespread. When considered in a continuum of assessments based on the proximity to instruction, benchmark assessments are located between teachers' minute-by-minute and daily formative assessment practices that are used to direct instruction to support learning, and the summative unit assessments, or tests administered after instruction has occurred to measure learning. As such, the intended purpose of benchmark assessments blends the ideas of data-driven decision making with the principles of formative assessment. The expectation is that school administrators and teachers will use these test results to identify students' misunderstandings and correct the course of learning in preparation for the year-end state mandated exams. Examining the extent to which benchmark assessments results are being used in this formative way was the primary aim of this study. This report presents results of a survey of elementary and middle school teachers in four school divisions about their use of benchmark assessment data to improve instruction and support student learning. This report documents the second phase of a two-stage investigation of teachers' formative uses of benchmark assessment results.

Literature Review

Benchmark Assessments as Formative Assessment

The enactment of No Child Left Behind (2001) dramatically increased pressure on schools to raise student achievement and address achievement gaps. In response, school districts have developed and implemented interim or benchmark assessments to provide data that can be used throughout the school year to monitor student achievement and progress toward meeting

3

state curricular standards as teachers prepare students for the end of the year state mandated tests. Perie, Marion, Gong and Wurtzel (2007) offered a definition of benchmark assessments relative to the purpose of formative and summative assessments. Perie et al. defined interim or benchmark assessments as those that "(1) assess students' knowledge and skills relative to curriculum goals within a limited time frame, and (2) are designed to inform teachers' instructional decisions as well as decisions beyond classroom levels" (p. 4). Benchmark assessments provide data that can be used at the classroom or individual student levels in addition to aggregated across classrooms and/or schools.

The potential for use of benchmark assessment to guide teaching practice and inform instructional adjustments is what makes these types of assessment characteristically formative. The specific practices that constitute formative assessment are highly varied in the literature, so much so that Bennett concluded "the term formative assessment does not yet represent a well-defined set of art[i]facts or practices. A meaningful definition requires a theory of action…" (2011, p. 19). Earlier work by Black and William (2009) addressed this issue and put forth a theory of formative assessment. They provide a working conceptual framework of action for formative assessment that includes five essential strategies:

1. Clarifying and sharing learning intentions and criteria for success;
2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding;
3. Providing feedback that moves learners forward;
4. Activating students as instructional resources for one another; and
5. Activating students as the owners of their own learning (p. 8).

This framework identifies the ways in which benchmark assessment data can serve formative functions and guide decisions about how to implement specific activities associated with formative assessment. The prevalence of benchmark assessments has grown significantly over the last several years. In a synthesis of four separate studies related to data-driven decision making, Marsh, Pane and Hamilton (2006) reported 89% of school districts in Georgia required some or all schools to administer benchmark or "progress" tests in mathematics; 50% required similar tests in science. One-half of California districts and one-third of districts in Pennsylvania required benchmark assessments in mathematics. In a more recent survey of urban school districts, Burch (2010) described 82% reported having implemented some form of benchmark assessment, and of these, 69% had begun implementation following the enactment of NCLB. Even though these data suggest the extensive use of benchmark assessments, Marsh et al. concluded that little is known about how these tests are influencing instruction, and as a result, student achievement (2006). Consequently, teachers' use of interim or benchmark assessment data to engage in formative assessment practices with the specific goal of improving student achievement has become a growing area of interest for school districts and the assessment research community.

Teachers' Approaches to Benchmark Assessment Data Analysis

Given the more recent interest in teachers' use of benchmark assessment data, the literature in this area is relatively sparse. There are however, a few studies that have investigated this topic. To understand how teachers were using interim assessment data, Oláh, Lawrence, and Riggan (2010) conducted extensive interviews, coupled with a data analysis scenario, with a sample of 25 teachers in five different elementary schools in the School District of Philadelphia. In addition to interviews, they collected relevant documents, such as copies of the grade 3-5

5

mathematics assessments, classroom assessments, student work samples and teacher-developed templates used to organize assessment data. Oláh et al. found that teachers analyzed data in two main ways. The large majority of teachers (86%) first used the data to locate or identify errors based on correct or incorrect responses. After identifying the items that were answered incorrectly, teachers then used the errors to diagnose why students may have selected the wrong response (Oláh et al., 2010; Riggan & Oláh, 2011). For the majority of teachers, initial steps in analysis involved linking student weakness with content standards; few teachers began their analysis with examining the data to identify poor performing individual students.

Based on their findings, Oláh et al. were able to establish a common analytical framework where (1) teachers identified student weaknesses as indicated by item-analysis related to content and specific students; (2) teachers engaged in a validation process to ensure that items and responses were accurate indicators of students' understanding of mathematics concepts and skills; (3) once the information was considered valid, teachers established a "context for interpretation" where the assessment data was compared to teachers' own standards for student performance; and (4) teachers developed an instructional response based on their analysis. Within this framework, teachers were clearly making connections between student weaknesses, as indicated by incorrect responses, and the content of instruction.

Blanc, Christman, Liu, Mitchell, Travers, and Bulkley (2010) described a slightly different pattern of how teachers focused their analysis of interim assessment data based on their work in the Philadelphia school district in 2006-2007. Blanc et al. relied on multi-method sources of data, including a district-wide teacher survey and extensive interviews with school administrators and teachers, as well as observations of team and grade group meetings.   They found teachers used data to (1) identify students on the bubble of moving from one proficiency

6

category to the next (e.g., basic to proficient or below basic to basic) and to deliver specific interventions to improve academic performance; (2) identify the content and skills that need to be re-taught; (3) identify students who have similar misunderstandings or skill deficits who could be grouped to provide tailored instruction; (4) evaluate classroom routines and make adjustments to enhance motivation and engage students in taking responsibility for learning; and (5) identify content and instructional needs to inform professional development opportunities and other supports required to strengthen teachers' skill level with data use.

Another more recent study highlights how teachers use different types of data at item, individual student, and classroom levels to shape their instructional responses. Similar to Oláh et al. (2010), Shepard, Davidson, and Bowman (2011) focused their research on teachers' use of mathematics benchmark assessment data, at the middle rather than elementary school level. Shepard et al. conducted two interviews with each of the 30 teacher participants, representing seven different school districts. Teachers were selected from schools that were identified as effectively implementing the district assessment. Similar to Oláh et al., Shepard et al. found that teachers described students' mastery of content as the primary source of information gained from the benchmark assessment results. Teachers typically described mastery of content according to different levels of specificity, including "broad-progress information; standards-focused information combined with item-level information; and primarily item-level information" (p. 14). Teachers also described using the assessment data to evaluate or examine their own instruction. When prompted about the specific insights gained from the assessment information, Shepard et al. noted that less than half of the participants were able to describe any insights in depth. A few teachers described procedural insights while a greater number noted gaining information about specific test-taking skills. The level of generality with which teachers

described the information they acquired from the assessment score reports led Shepard et al. to conclude that the data provided by benchmark assessments were not sufficient to direct teachers' instructional responses other than to re-teach weak content. This finding, compared to those of Oláh et al. and Blanc et al. (2010), suggests that the way in which data are reported and organizational factors such as, school leadership and supports for data use, may substantially influence the extent to which teachers use data to inform instruction.

Instructional Uses of Benchmark Assessment Data

With regard to the instructional use of benchmark assessment data, the findings of several studies indicate that teachers are using results to make instructional adjustments, such as identifying and addressing areas of student weakness, providing remediation for gaps in student learning, setting instructional priorities and increasing efficiency, determining instructional approaches, and differentiating instruction for small groups or customizing learning activities for individual students (Brunner et al., 2005; Christman et al., 2009; Marsh et al., 2006; Oláh et al., 2010; Shepard et al., 2001; Yeh, 2006). The literature indicates that teachers have three primary instructional responses, depending on the scope of student misunderstandings as suggested by assessment data: (1) providing remediation for individual or smaller groups of students; (2) re-teaching, which typically focuses on providing additional instruction using a different strategy to the class as a whole; and (3) grouping students. A descriptive study of 45 elementary teachers, using interviews, observations and surveys, found that benchmark assessment data "did not substantially change their instructional and assessment practice" (Goertz et al., 2009, p. 6). They found that benchmark data influenced what was taught, but not how to re-teach. Similarly, Christman et al. reported that school leaders and teachers were not maximizing the potential of

benchmark assessment data to provide for deeper conversations about instructional content and learning processes.

More recent studies seem to have recognized extant findings about the lack of specificity or the generality of data-based instructional responses, and have been designed to hone in on the direct links teachers are making between data analysis and their own efforts to promote learning. For example, Oláh et al. incorporated a "data analysis scenario" into a series of teacher interviews to pinpoint their thinking about assessment data and its relevance to teaching. They described how teachers used data to "diagnose" students' misconceptions and understandings, and found that, by and large, teachers focused on procedural aspects of student errors and attributed conceptual misunderstandings to external factors or other cognitive difficulties. Like other studies, teachers used data to focus on re-teaching at the classroom level or to small groups depending on the extent of the misunderstandings. They were less likely to describe remediation practices. Oláh et al. indicated that teachers tended to emphasize procedural steps or processes in their re-teaching which may or may not have involved the use of new or different strategies. According to Oláh et al., teachers are using assessment results but are not necessarily making strong associations between students' conceptual misunderstandings and an appropriate instructional response. This may be due to limited conceptual information that can be obtained from the assessments.   What is clear from the literature is that teachers are using assessment data to identify students' strengths and weaknesses and are comfortable linking student test results to content standards (Blanc, et al., 2010; Oláh et al., 2010; Shepard et al., 2011). What is less clear are the pedagogical connections teachers are making between their re-teaching efforts and the nature of students' misunderstandings.

Factors that Affect Data Use

In addition to the impact of benchmark testing policies on instruction and student outcomes, the literature suggests that a variety of factors are associated with teachers' formative use of benchmark test results. According to the literature, teachers' accessibility to the test results and their perceptions of data quality are two primary factors that influence their use of data. The timeliness and type of information teachers receive are viewed as critical to the extent to which test results can be considered "actionable information" or information on which to base educational decisions. For example, online access to data was associated with teachers' use of data (Marsh et al., 2006). The RAND synthesis also suggested that teachers reported concerns about the reliability and validity of test scores, especially when they perceived a lack of alignment of the tests with the curriculum, and when they were concerned about students' trivial attitudes toward the test (Marsh et al., 2006). Other studies have pointed to the need to provide capacity and professional development for teachers to support their use of benchmark testing data (Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Murnane, Sharkey & Boudet, 2005; Symonds, 2004; Trimble, Gay, & Matthews, 2005; Vogel, Rau, Baker & Ashby, 2006; Wohlstetter, Datnow & Park, 2008).

Research related to building capacity for data use describes the essential role of effective leadership and school administrators in developing a data-driven decision making culture and system. Halverson et al. (2005) describes how the role of school administrators has evolved to include "creating accountable learning systems in schools" (p. 5). Supovitz and Klein (2003) found that schools using data in innovative ways also had strong visionary leadership. Principals set the expectations for faculty and staff and in doing so can create supportive environments in which to address student learning through the use of assessment data. Copland (2002) found that

10

school leaders were most effective in building and supporting a culture for data use if they adopted distributed leadership approaches and involved teachers on a broad scale. In addition to effective leadership, time is also an essential ingredient for teachers' use of data. This includes time for professional development as well as time set aside to analyze and discuss assessment data with colleagues (Goertz et al., 2009). Research suggests that time to collaborate or engage with colleagues in professional learning communities can support teachers' effective use of data. However, the extent to which time during the school day is allocated to teachers' use of data is limited. Based on their 2007 national survey of K-12 teachers, Means, Padilla, DeBarger, Bakia (2009) reported that 23% of respondents have time during the school day to analyze data, while 59% reported needing to access data outside of the regular work day.

In addition to organizational structures such as distributed leadership, professional development, and scheduled time for collaboration, teacher characteristics also influence the extent to which benchmark assessment data can be used effectively. Several studies cite teachers' lack of expertise in analyzing and interpreting test score information and the need to develop a level of assessment literacy to support the effective and meaningful use of test score information (Kerr et al., 2006; Murnane, Sharkey & Boudet, 2005; Symonds, 2004; Trimble, Gay, & Matthews, 2005; Vogel et al., 2006; Wohlstetter, Datnow & Park, 2008). Other research indicates that teachers with strong content knowledge are more flexible and can easily adapt instruction to meet students' learning needs akin to formative assessment processes (Duschl & Gitomer, 1997; Fennema, Franke, Carpenter, & Carey, 1993). Strong content knowledge enables teachers to target students' conceptual understandings of the instructional content. Goertz et al. (2009) found teachers who were focused on students' conceptual understanding were more likely to craft instructional responses based on assessment data rather than making organizational

responses such as using data for grouping students. Similarly, Datnow, Park, and Wohlstetter (2007) contend teachers' capacity and ability to use assessment data is deeply tied to their instructional knowledge. Analysis of data helps teachers to identify student learning problems, but does not necessarily direct teachers toward a specific instructional solution. These studies point to the foundational knowledge and skills necessary to build teachers' capacity for making connections between students' conceptual misunderstandings and instruction.

Impact on Student Achievement

The logic behind the implementation of benchmark assessments is relatively straightforward – the tests provide principals and teachers with periodic information about student progress; assessment results can be used to identify students' strengths and weaknesses and subsequently modify instruction to enhance student learning. Empirical evidence that the use of benchmark assessment data has had a positive impact on student learning is both limited and mixed. For example, some research suggests that targeted instruction can lead to improvements in student test scores (Lachat & Smith, 2005; Nelson & Eddy, 2008; Trimble, Gay & Matthews, 2005; Yeh, 2006) as well as proficiency in reading and mathematics (Peterson, 2007). However, empirical investigations based on quasi-experimental designs have found no significant differences between schools using benchmark assessments and comparison schools not using such tests (Henderson, Petrosino & Guckenburg, 2008; Niemi, Wang, Wang, Vallone, & Griffin, 2007). Other studies suggest benchmark testing can lead to positive impacts on factors that may ultimately contribute to improved student achievement, such as increased student engagement and motivation (Christman et al., 2009; Yeh, 2006), and greater access to learning opportunities including tutorial and remediation instruction or services (Marsh et al., 2006).

12

Some of the lackluster findings about the use of assessment data may be influenced by long-held expectations for the potential impact of formative assessment practices on student achievement. Black and Wiliam's 1998 seminal work is widely regarded as evidence of the positive effect formative assessment can have on student achievement. Based on their synthesis of studies of formative assessment, they concluded that typical effects of these studies were between .40 and .70 and that these effects were larger than those of most educational interventions. More recently, researchers have identified methodological limitations and concerns about the validity of conclusions drawn from Black and Wiliam's 1998 review (Dunn & Mulvenon, 2009; Kingston & Nash, 2011). Kingston and Nash identified several limitations of the early work conducted by Black and Wiliam. They conducted their own meta-analysis of studies using formative assessment to determine the average effect size of formative assessment on student achievement, while accounting for previous study limitations. Based on a sample of 42 studies, they found a weighted mean effect size of .20 with a median of .25; results substantially lower than earlier estimates of the effects of formative assessment. However, Kingston and Nash noted that even with the lower effects, formative assessment can still provide for improved student learning (2009). Given these findings, formative assessment practices are effective ways to support student learning. When teachers' decisions about formative assessment are informed by student achievement data, and then targeted toward students' learning needs, there is a clear potential to improve learning. Consequently, the purpose of this study was to explore how teachers report using benchmark assessment results in formative ways.

Methodology

Based on the first qualitative phase of our research on teachers' use of benchmark assessment data (Abrams, Wetzel, & McMillan, 2010) we found that teachers approached their analysis of benchmark assessment results first by examining the data to identify content missed by large groups of students and would then link these items to specific content standards on which to base their instructional response. Teachers also described analyzing data to identify individual students who performed poorly to determine who needed more individualized instruction in greater depth – this often translated to more time spent with students one-on-one. When asked about the specific ways teachers adjusted their instruction, they described providing additional homework, weaving short reviews into class instruction in the form of questions in warm-up activities, or providing workbook exercises or worksheets. Similar to Goertz et al. (2009) and Shepard et al. (2011), teachers provided limited discussion of their specific modifications to the delivery of content or instructional strategies. With this finding in mind, we attempted to target teachers' instructional responses to benchmark assessment data in this follow-up quantitative study and posed the following research questions:

1. What conditions are necessary to promote use of benchmark assessment results?

2. How do teachers report analyzing and using benchmark assessment results to inform instruction? To inform decisions about students?

3. What factors most influence teachers' use of benchmark assessment results?

Study Design and Instrumentation

To address the research questions a survey research design was implemented. A 40-item survey was developed to measure a variety of topics related to benchmark assessments. The development of survey items was informed by the findings of our earlier qualitative work, that

14

was conducted with teachers in the target school districts, as well as an instrument developed by

the American Institutes for Research and the Council of Great City Schools for their *Urban Data

Study: The Use of Interim Assessment Data in Urban Schools: Links among Data Use Practices

and Student Achievement* (report forthcoming).   The final survey included multiple selected-

response questions related to: (1) district and school testing policies, (2) teachers' access to

benchmark assessment results, (3) how teachers analyze assessment results, (4) instructional uses

of results, (5) general attitudes and opinions toward benchmark testing, and (6) demographic and

individual characteristics. Two open-ended questions were included at the end of the survey that

asked teachers to "describe a situation or instance in which benchmark test results were

especially useful in making decisions about your teaching or making decisions about students."

The majority of the selected-response items required respondents to select from Likert-scale

options. See Appendix A for the complete survey instrument. Prior to implementation, the

survey was piloted with elementary school teachers from a local school division that did not

intend to participate in the larger-scale study. Participants in the pilot administration were asked

to review the survey for the clarity of the directions, questions, and response scales and provide

their opinions about ways to improve the formatting of the measure for ease of administration.

Survey Administration and Response

The survey was administered electronically using *Inquisite Survey Software* (version 9).

Elementary (grades 4 and 5) and middle school teachers of core-content areas were the target

population. A link to the survey was sent by electronic mail directly to teachers by school district

personnel along with an email message from the research team. Teachers in four school districts

surrounding a southeastern urban area were surveyed in February and March 2011 following the

administration of the 3rd quarter benchmark assessments. These schools districts were selected

for participation in the study because they all shared membership in a local research consortium and had identified benchmark assessments as a common priority. All of the districts were interested in understanding how benchmark assessment policies were influencing instruction and how they could best support teachers' use of the data. All participating districts had begun implementation benchmark assessments during the same school year (2007-2008). A total of 460 teachers responded to the survey, of these 390 provided usable responses. The response rate across the four participating school districts varied, and ranged from 25% to 85%. It was not possible to calculate the response rate for one of the districts due to modifications in the teacher recruitment procedures for this district.

Responding Teacher Characteristics

As shown in Table 1, 43% of participating teachers taught in the elementary grades and the remaining 57% percent were in middle schools. At the middle school level, the majority of respondents taught either Reading/English Language Arts (31%) or Mathematics (29%). Roughly 20% of responding middle school teachers taught Science and Social Studies respectively. The vast majority were female (84%) and held Master's degrees (65%). On average, teachers had been in the classroom for 14.5 years, with the majority (60%) having 11 or more years of teaching experience. Teachers also reported working in their current school for 8 years on average.

Table 1

Characteristics of Teacher Respondents

Characteristics	*n*	*%*
Gender		
Males	61	15.7
Females	328	84.3
Race		
Hispanic/ Latino	2	0.5
White	362	93
Black/ African American	21	5.4
Asian	1	0.3
Native Hawaiian or other Pacific Islander	0	0.0
American Indian or Alaska Native	1	0.0
Other	3	0.8
Grade Level		
Elementary	169	43.3
Middle	221	56.7
Teaching Experience		
0- 5 years	56	14.9
6- 10 years	93	24.8
11 + years	226	60.3
Grades Teaching		
Fourth	85	21.8
Fifth	97	24.9
Sixth	78	20.0
Seventh	85	21.8
Eighth	95	24.4
Other	6	1.5
Subjects Teaching		
All (Elementary)	102	26.2
Reading/ English Language Arts	119	30.5
Mathematics	111	28.5
Science	76	19.5
Social Studies	78	20.0
Educational Qualification		
Bachelor's Degree	382	97.9
Master's Degree	197	65.0
Educational Specialist/ Professional Diploma	36	18.8
Certificate of Advanced Graduate Studies	21	11.7
Doctoral or Professional Degree	0	0.0

Note. Total sample size, *N*=390. Subtotals may not add to 390 on account of missing data.

Data Analysis

The survey data were analyzed using two main approaches to examine teachers' reported use of benchmark assessment data. Initial analyses included basic descriptive statistics and measures of variability to report on trends and patterns in the data. Following these analyses, Pearson correlations were computed to identify relationships and strong associations among different organizational or contextual conditions and teachers' reported use of benchmark assessment data for instruction.   These associations were further examined using linear regression analyses, to identify those conditions most predictive of teachers' use of benchmark test data.

The reporting of the survey results is divided into two sections. The first, the "descriptive results", is organized into four sections: (1) review and analysis of benchmark data; (2) influences on instruction; (3) use of results; and (4) general attitudes about benchmarks. Within each section, teachers' percentage responses to individual survey items are reported and discussed. The second section, "factors influencing teachers' data use", reports on the relationships found among variables constructed from the item-level survey data. This section reports on the associations among district and school-level organizational conditions and teachers' instructional use of benchmark assessment results.

Descriptive Results

The following section reports teachers' responses to the survey items using descriptive results (e.g., percentages, mean, and standard deviation) and is organized according to the major topics measured in the survey.

18

Teacher review and analysis of benchmark test results. A main section of the survey focused on how teachers analyzed benchmark results. Questions in this section addressed how teachers analyzed results, how often, and in what context their review occurred.

Accessing results. Several survey questions asked teachers about their access to benchmark assessment results. A large majority of teachers (78%) reported receiving results immediately after the administration of benchmark assessments, and about 22% reported receiving the results at least 24 hours after the test administration. As shown in Table 2, most teachers (87%) reported receiving the results electronically. On whether assessment questions were provided with the results, about 40% of teachers reported they did not receive the test questions with the results.

Table 2

Access to Benchmark Assessment Results

Items	*n*	*%*
Are the assessment questions provided with the results?		
Yes	241	62.1
No	147	37.9
How long after the administration of the most recent benchmark assessments were results made available to you?		
Immediate (within 24 hours)	301	78.2
Delayed (after at least 24 hours)	84	21.8

Frequency of review. Table 3 shows how frequently teachers reviewed data with other teachers, school administrators, students and parents. In general, most teachers reviewed benchmark assessment results with others about 1-2 times a quarter; including students (67.6%), school administrators (64%), other grade level teachers (58%), department or grade-level chair (55%), parents (54%), grade-level lead teacher (45%), instructional coaches (29%) and division central office staff (16%). As shown in Table 3, some teachers reported reviewing results more

frequently, about 1-2 times a month with other grade-level teachers (19%), their grade-level lead

teacher (15%), and their grade-level chair (13%), students (12%), administrators (11%),

instructional coaches (11%), and parents (9%).

Table 3

Frequency and Context of Teachers Review and Analysis of Benchmark Assessment Results

Items	Never		1-2 times a quarter		1-2 times a month		1-2 times a week		*M*	*SD*
	n	*%*	*n*	*%*	*n*	*%*	*n*	*%*		
Department chair/grade-level chair	92	28.3	179	55.1	41	12.6	13	4.0	1.81	0.76
Grade level lead teacher	96	31.7	135	44.6	46	15.2	26	8.6	1.86	0.92
Other classroom teachers in my grade level or subject area	31	8.6	209	58.2	67	18.7	52	14.5	2.26	0.85
Instructional coaches	141	56.6	72	28.9	28	11.2	8	3.2	1.54	0.77
School administrators	81	23.5	221	64.2	39	11.3	3	0.9	1.84	0.59
Division central office staff	241	82.0	46	15.6	7	2.4	0	0.0	1.19	0.46
Parents/guardians	113	34.3	177	53.8	30	9.1	9	2.7	1.72	0.72
Students	22	6.2	240	67.6	44	12.4	49	13.8	2.27	0.78

A smaller percentage of teachers also reported reviewing results 1-2 times a week with other

teachers in the same grade-level or subject area (15%), students (14%) and the grade-level lead

teacher (9%). A majority of teachers reported that they never reviewed benchmark assessment

results with district central office staff (82%), and instructional coaches (57%), while many

reported they never discussed benchmark results with parents (34%), their grade-level lead

teacher (32%), school administrators (24%), or their grade-level chair (28%).

Time spent reviewing results. As reported previously, most teachers reviewed

benchmark assessments with others about 1-2 times a quarter, and more frequently with school

administrators, grade-level colleagues and students. Teachers were asked how much time they

spent reviewing results of the most recent benchmark administration. As shown in Table 4,

almost all teachers reported reviewing results independently with about 71% spending less than

two hours, and the remaining teachers devoting two or more hours to analysis.

Table 4

Time Spent Reviewing and Analyzing Benchmark Assessment Results

Items	0 hours		<1 hour		1-2 hours		2-3 hours		More than 3 hours		*M*	*SD*
	n	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*		
Independently	8	2.2	101	27.9	155	42.8	53	14.6	45	12.4	3.07	1.00
With teachers at my grade level	39	10.7	159	43.8	118	32.5	35	9.6	12	3.3	2.51	093
With teachers in other grade levels	251	69.3	86	23.8	19	5.2	4	1.1	2	0.6	1.40	0.69
With principal	200	55.2	124	34.3	28	7.7	8	2.2	2	0.6	1.59	0.77
With assistant principal	212	58.9	109	30.3	32	8.9	6	1.7	1	0.3	1.54	0.75
With content area coach	247	68.6	64	17.8	.7	10.3	9	2.5	3	0.8	3.11	1.34
With a data coach	317	88.3	16	4.5	17	4.7	7	1.9	2	0.6	1.22	0.67
With students	40	11.1	158	43.9	109	30.3	36	10.0	17	4.7	2.53	0.98
With parents	179	49.4	154	42.5	22	6.1	5	1.4	2	0.6	1.61	0.71
Other	226	93.4	4	1.7	4	1.7	3	1.2	5	2.1	1.17	0.71

Many teachers also reported spending between less than two hours reviewing results with

other grade-level teachers (76%), students (75%) parents (49%), and the school principal (35%).

About a third of teachers spent less than an hour discussing the results with the principal or

assistant principal. While a strong majority of teachers reported spending no time discussing

results with a content-area coach (69%) or a data coach (88%), a small percentage did engage in

these types of discussions.

Usefulness of review and analysis of benchmark test results. Teachers were also asked about the usefulness of reviewing benchmark data independently and with colleagues. Among the individuals with whom teachers reviewed data, most teachers found reviewing data with grade-level teachers (77%) and students (75%) to be somewhat or very useful. Teachers were divided about the usefulness of reviewing data with parents; close to 50% reported that it was somewhat or very useful, and the rest reported it was not at all or not very useful. A majority of teachers reported that reviewing results with the principal (64%), assistant principal (65%), content-area coach (70%) and data coach (86%) was not at all or not very useful.

Context. In addition to frequency and time, teachers were also asked about the context in which they used benchmark test data in their interactions with colleagues (see Table 5). Many teachers reported discussing student work (45%) and meeting with grade-level teams to examine trends in the data (40%). As shown in Table 5, item means suggest that for most teachers they engaged with colleagues about benchmark data to examine student progress or teaching practice from a slight to a moderate extent. In other words, these types of interactions were not a major component of the responding teachers' data use practices.

Influences on instruction. Another main area addressed in the survey included teachers' instructional uses of benchmark results. Teachers were asked to report on several different factors known to influence classroom instruction, specific aspects of benchmark data that were helpful in understanding student performance, as well as the usefulness of specific types of test data.

Policies and assessments. Most teachers reported that the state's curriculum framework and content standards (88%), the division's curriculum framework (83%), their own classroom observations (81%), and division pacing guides (67%) were major influences on the content and

focus of their instruction (see Table 6). Teachers also reported that their own teacher-developed classroom assessments (96%), local division policies and initiatives (87%), end-of-year assessment scores (82%), and curriculum-based unit assessments (72%) had a major influence on their classroom practice. When asked specifically about the influence of benchmark test data on their instruction, close to 48% of teachers reported that the results had a moderate influence compared to similar percentages that reported the data had a major (24%) and a minor (23%) influence on their instruction.

Table 5

Teachers Interactions based on Benchmark Assessment Results

Context	Not at all		Slight Extent		Moderate Extent		Major Extent		*M*	*SD*
	n	*%*	*n*	*%*	*n*	*%*	*n*	*%*		
Meet with grade-level teams or department teams to look at trends in the data (or analyze data)	35	9.7	125	34.6	144	39.9	57	15.8	2.62	0.87
Share ideas about using data to improve teaching with other teachers	45	12.5	113	31.3	137	38	66	18.3	2.62	0.92
Share and discuss student work with other teachers	23	6.4	98	27.1	164	45.4	76	21.1	2.81	0.84
Discuss particular lessons that were not very successful	41	11.4	119	33.0	142	39.3	59	16.3	2.61	0.89

Table 6

Factors that Influence Classroom Instruction

Item	No Influence		Minor Influence		Moderate Influence		Major Influence		*M*	*SD*
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*		
State's curriculum framework or content standards	1	0.3	3	0.9	36	10.9	290	87.9	3.86	0.39
Division's curriculum framework, standards, or guidelines	6	1.8	10	3.1	34	10.4	276	84.7	3.78	0.59
Division policies and initiatives	8	2.5	31	9.6	93	28.7	192	59.3	3.45	0.77
Benchmark assessment data	17	5.2	76	23	158	47.9	79	23.9	2.91	0.82
Assessments that you develop	6	1.8	7	2.1	129	39.6	184	56.4	3.51	0.64
Curriculum-based unit assessments	26	8.4	43	14	127	41.2	112	36.4	3.06	0.92
End-of-year state assessment scores	10	3.1	38	11.9	104	32.6	167	52.4	3.34	0.81
Division pacing guides	6	1.8	25	7.6	74	22.6	223	68	3.57	0.74
Your own classroom observations	3	0.9	8	2.4	50	15.2	267	81.4	3.77	0.53

Types of Data. Also related to factors influencing instruction were benchmark test results

reports. When asked what information was most helpful or influential for making instructional

decisions, 65-70% of teachers reported that assessment results on individual student performance

and the number/percentage of students that correctly answered each test item were the most

helpful. By comparison, 25-30% of respondents found this information to be somewhat helpful.

Teachers were asked how much they used specific statistics provided by benchmark

assessment reports such as the percentage of students scoring above the proficient level, results

for student subgroups, aggregated forms of results by class and grade levels, and results for

individual test items. A large majority of teachers (71-91%) reported moderate to extensive use

of results that showed student performance on individual test items, summary class-level results and the percent of students scoring at or above the proficient level.

Teachers' use of results.

Inform decisions about individual students and instruction. Several survey questions asked teachers how they used benchmark test results to inform decisions about individual students in their classroom. Most teachers (82%) reported moderate or extensive use of benchmark test data to identify students in need of remedial assistance (see Table 7). Similarly, teachers reported moderate (45%) to extensive (36%) use of results to identify and correct gaps in the curriculum, as well as tailor instruction to the needs of individual students (41% and 31% respectively). Teachers were least likely to use benchmark test results to develop student Individual Educational Plans (IEPs).

Teachers varied in their responses to questions about the extent to which they involved students in the use of benchmark data. Many of them used data 3-4 times a year to inform students (58%) and parents (42%) of their progress. However, a sizable percentage (35-43%) of teachers reported never involving students in their interpretations of student performance or used results to create new strategies for learning.

Teachers were also asked more targeted questions about their specific use of benchmark test data to help address students' instructional needs. As shown in Table 8, many teachers reported that they reviewed key concepts with the entire class as a result of benchmark assessment scores to a major extent (43%) or moderate extent (42%). Between 38-42% of teachers reported changing the sequence of instruction, modifying the skills taught and using individualized instructional approaches during class to help support student learning at least to a moderate extent based on benchmark assessment results.

Table 7

Use of Benchmark Assessment Results for Decisions about Individual Students

Item	Did not use in this way		Minimal Use		Moderate Use		Extensive Use		*M*	*SD*
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*		
Identify individual students who need remedial assistance	20	5.5	46	12.7	145	40.1	151	41.7	3.18	0.86
Diagnose learning problems	99	27.4	110	30.5	114	31.6	38	10.5	2.25	0.97
Tailor instruction to individual students' needs	32	3.6	68	15.5	163	45.2	129	35.7	2.94	0.94
Identify and correct gaps in the curriculum for all students	13	3.6	56	15.5	163	45.2	129	35.7	3.13	0.81
Recommend tutoring or other educational services for students	82	22.8	86	24.0	123	34.3	68	18.9	2.49	1.04
Identify areas where I need to strengthen my content knowledge or teaching skills	30	8.3	66	18.3	158	43.9	106	29.4	2.94	0.90
Assign or reassign students to classes or groups	155	43.3	68	19.0	88	24.6	47	13.1	2.08	1.10
Determine instructional materials to use with my class(es)	63	17.5	82	22.8	132	36.8	82	22.8	2.65	1.02
Develop or revise Individualized Education Programs (IEPs)	204	57.5	78	22.0	47	13.2	26	7.3	1.70	0.96

Table 8

Use of Benchmark Assessment Results for Decisions about Instruction

Item	Not at all		Minor Extent		Moderate Extent		Major Extent		*M*	*SD*
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*		
Reviewed key concepts for the entire class	11	3.4	39	11.9	138	42.2	139	42.5	3.24	0.79
Used same-level achievement groupings	106	32.7	80	24.7	107	33.0	31	9.6	2.19	1.00
Used mixed-level achievement groupings	97	29.8	83	25.5	105	32.3	40	12.3	2.27	1.02
Used individualized instruction during class to address the needs of struggling students	37	11.3	81	24.8	124	38.0	84	25.8	2.78	0.96
Provided individual assistance outside of class to address the needs of struggling students	60	18.4	90	27.6	101	31.0	75	23	2.59	1.04
Changed the sequence of instruction	90	27.9	87	26.9	107	33.1	39	12.1	2.29	1.01
Added, deleted, or changed skills taught	50	15.4	92	28.3	134	41.2	49	15.1	2.56	0.93
Changed teaching method (e.g. lecture, cooperative learning, student inquiry)	54	16.5	87	26.6	130	39.8	56	17.1	2.57	0.96

Influence instructional practice. In addition to asking teachers about the extent of their

use of benchmark assessment results, the survey also included items that measured the *influence*

of the results on decision-making. As shown in Table 9, teachers reported benchmarks results to

have a moderate influence on identifying students who needed tutoring or supplemental

instruction (43%), adjusting pacing (37%) and adjusting goals for student learning (34%).

According to teachers' responses, the benchmark assessment results had the least influence on

decisions about retaining students in the same grade level, grouping students for instruction, or

adjusting textbooks and teaching materials, with 60-79% of teachers reporting minimal or no

influence at all in these categories. Similarly, many teachers reported benchmark results had no or minimal influence on teachers' professional evaluations (75%) or identifying professional development needs (68%). These data suggest that other factors, in addition to benchmark assessment results, may contribute to informing decisions in these areas.

Table 9

Influence of Benchmark Assessment Results on Decisions about Instruction and Teachers

Item	No Influence		Minor Influence		Moderate Influence		Major Influence		*M*	*SD*
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*		
Determining a student's grouping for instruction	100	30.8	97	29.8	94	28.9	34	10.5	2.19	0.99
Adjusting goals for student learning	66	10.3	105	32.3	111	34.2	43	13.2	2.40	0.96
Adjusting pacing in areas where students encountered problems	54	16.7	95	29.3	119	36.7	56	17.3	2.55	0.96
Adjusting use of textbooks and instructional materials	102	31.6	104	32.2	91	28.2	26	8.0	2.13	0.95
Identifying students to be retained at the same grade level	171	52.9	84	26.0	57	17.6	11	3.4	1.72	0.87
Identifying students for tutoring or other supplemental instruction	43	13.3	72	22.2	140	43.2	69	21.3	2.73	0.95
Identifying professional development needs	117	36.1	104	32.1	85	26.2	18	5.6	2.01	0.92
Evaluating teachers	157	49.1	83	25.9	55	17.2	25	7.8	1.84	0.98

Adjust instruction. Teachers were also asked to report on the extent to which they made changes to their teaching content and strategies, expectations and assessment practices based on the results of the most recent district benchmark assessment administration (see Table 10). Between 39-42% of teachers reported they made no changes to their teaching content, or their expectations for student performance based on benchmark results. Responses to survey questions asking if teachers changed the instructional strategies were almost evenly split, with 35%

28

reporting they made only minor changes compared to 35% that had made moderate changes. Similar responses were shown for questions about their own classroom assessments. When asked if they changed the mix of assessments used to evaluate students, 32% reported making minor changes compared to 32% reporting moderate changes, and 26% reporting no changes. Teachers were also asked about the extent to which they increased or decreased the use of specific instructional practices based on students' benchmark test performance. A sizable percentage (40%) reported increasing cooperative learning and group work, and almost 50% reported using more problem-solving activities in their classroom. When presented with a variety of instructional practices including textbook assignments, use of worksheets, portfolios, lecturing, and writing assignments, for example, many teachers (42%) reported making no instructional changes to the content of their instruction or the teaching strategies they used (18%) on the basis of benchmark assessment results.

Table 10

Extent of Instructional Change Determined by Benchmark Assessment Results

Item	Not at all		Minor Extent		Moderate Extent		Major Extent		*M*	*SD*
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*		
The curriculum content I teach	137	42.4	102	31.6	68	21.1	16	5.0	1.89	0.91
My expectations for student performance	127	39.3	85	26.3	94	29.1	17	5.3	2.00	0.95
The instructional strategies I employ	59	18.3	113	35.0	114	35.3	37	11.5	2.40	0.92
The types of mix of assessments I use to evaluate students	85	26.2	105	32.4	104	32.1	30	9.3	2.24	0.95

Proficiency with using benchmark data. In order to assess teachers' proficiency in using benchmark assessment data, they were asked to report on the level of their proficiency with

specific activities related to data use. A sizable majority of teachers (84-90%) felt moderately or very proficient in analyzing trends, interpreting student strengths and weaknesses, incorporating data into lesson planning, and adapting teaching based on benchmark assessment data. Many teachers reported feeling moderately proficient (45%) in analyzing data by Adequate Yearly Progress (AYP) student subgroups to improve learning and test performance.

Barriers to data use. Teachers were asked to report on the extent to which factors such as time, resources, professional development and data analysis skills influenced their ability to use data to make instructional decisions. As shown in Table 11, teachers were almost equally divided in their responses about the extent to which a lack of time to analyze data or discuss data with colleagues limited their ability to use benchmark results. A large majority of teachers (70-75%) reported that several factors did not limit their use of data including, personal discomfort with conducting data analysis, insufficient data, untimely reporting of results, or a lack of resources. In other words, the results suggest that there is sufficient personal and school capacity and information provided by the benchmark tests to enable teachers to use the results. The survey results also suggest that the primary barriers to data use are curriculum pacing pressures – with a majority of teachers (60%) reporting that these pressures affected their ability to use benchmark test results to a moderate or major extent.

General attitudes about benchmark testing. In addition to specific questions about teachers' analysis and instructional use of benchmark test results, the survey also included questions about general attitudes toward benchmark testing (see Tables 12 - 14). Teachers were asked to report on the extent to which benchmark assessments were aligned with division policies and standards. Overall, teachers agreed or strongly agreed that benchmark assessments

Table 11

Barriers to Benchmark Assessment Data Use

Item	Not at all		Minor Extent		Moderate Extent		Major Extent		*M*	*SD*
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*		
Lack of time to study and think about available data	72	22.4	85	26.4	85	26.4	80	24.8	2.51	1.09
Lack of time to collaborate with others in analyzing and interpreting data	73	22.5	84	25.9	80	24.7	87	26.9	2.54	1.11
Not enough professional development	150	46.3	89	27.5	64	19.8	21	6.5	1.86	0.95
Personal discomfort with data analysis	227	70.1	66	20.4	21	6.5	10	3.1	1.43	0.75
Lack of technology (e.g. access to computer with reliable internet connection)	229	70.9	51	15.8	28	8.7	15	4.6	1.46	0.83
Insufficient amount of data	242	74.9	59	18.3	17	5.3	5	1.5	1.33	0.65
Data provided too late for use	257	79.6	45	13.9	16	5.0	5	1.5	1.28	0.63
Curriculum pacing pressures	64	19.8	64	19.8	85	26.3	110	34.1	2.75	1.13
Division pacing guides do not allow me to re-teach based on results of benchmark data	97	30.3	64	20	64	20.0	95	29.7	2.49	1.21
Other	114	74	7	4.5	5	3.2	28	18.2	1.66	1.18

were well-aligned with state and school division standards (76%), state assessments (69%) and

pacing guides (75%). Teachers were mostly in agreement that benchmark assessments were

appropriately challenging for students (69%), and were well-aligned with the content of their

classroom instruction (78%). Many teachers (61%) disagreed or strongly disagreed that

benchmark assessments were of little use to instruction, suggesting that benchmark assessments

provide useful information for teachers.

Responses to survey items about attitudes toward school district support for data use indicated that many teachers (63%) recognized that there were clear and consistent goals for using data to support school improvement efforts. Teachers were divided in their agreement about whether district staff provided enough expertise and information to support data use at the school level. Similarly, teachers were divided about whether their school district data use policies helped to address students' needs. About 78% disagreed that their school district provided adequate resources to support their use of data.

Teachers also responded to questions about their access to benchmark results data. A majority of teachers agreed or strongly agreed (82-95%) that benchmark assessment results were provided in a timely manner and were easy to use and access. Many teachers (67%) also agreed or strongly agreed that there were enough computers to access benchmark data online. Roughly 75% of teachers reported that they used their personal time, rather than time during the school day, to access and review benchmark assessment data.

Responses to survey questions about teachers' perceptions of the professional climate in their school showed that 75-80% agreed or strongly agreed that the teachers in their school were continually learning and seeking new ideas, using student performance data, and engaging in inquiry and reflection. Approximately, 70% of respondents also agreed or strongly agreed that teachers in their school examined school-level performance data on assessments. A majority of teachers (55%) reported that the assessment of student performance led to curriculum changes. However, a smaller yet sizable percentage (36%) disagreed that curriculum changes were taking place.

Table 12

Teacher Attitudes about Benchmark Testing

Item	Strongly Disagree		Disagree		Agree		Strongly Agree		*M*	*SD*
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*		
Alignment of Benchmark Testing with Policy										
Well-aligned with the state assessment	25	8.2	67	22	164	53.8	49	16.1	2.87	.77
Well-aligned with the pacing guides	21	6.8	53	17.3	191	62.2	42	13.7	2.78	.81
Well-aligned with what I teach in the classroom	14	4.6	50	16.3	190	16.9	53	17.3	2.83	.75
Appropriately challenging for my students	26	8.5	67	21.9	171	55.9	42	13.7	2.75	.80
Division Policy and Support for Use of benchmark data										
The division sets clear, consistent goals for schools to use data for school improvement.	19	6.5	82	27.9	159	54.1	34	11.6	2.71	0.76
Division staff provides information and expertise that support the data use efforts at my school.	33	11.5	108	37.6	123	42.9	23	8.0	2.47	0.80
The division's data use policies help us address student needs at our school.	29	10.3	114	40.4	121	42.9	18	6.4	2.45	0.76
The division has designated adequate resources (time, staff, money) to facilitate teachers' use of data.	65	22.6	136	47.4	77	26.8	9	3.1	2.10	0.78

Table 13

Teacher Attitudes about Ease of Use of Benchmark Assessment Data

Item	Strongly Disagree		Disagree		Agree		Strongly Agree		*M*	*SD*
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*		
Benchmark assessment results are reported to me in a timely manner.	4	1.3	8	2.6	128	42.0	165	54.1	3.49	0.62
Benchmark assessment data are easy to use.	4	1.3	29	9.5	143	47.0	128	42.1	3.30	0.69
The division provides benchmark assessment data to schools in easy-to-use formats.	4	1.3	33	11.1	153	51.5	107	36.0	3.22	0.69
It is easy to access benchmark assessment data directly in the division data system.	14	4.9	27	9.4	135	46.9	112	38.9	3.20	0.80
My school's internet connection enables teachers to assess the division benchmark assessment system online.	14	4.9	19	6.6	131	45.5	124	43.1	3.27	0.79
There are enough computers at my school to enable teachers to access the division benchmark assessment system online.	30	10.2	27	9.2	138	46.9	99	33.7	3.04	0.92
If I want to use benchmark assessment data in my teaching, I have to use my personal time to review the data.	11	3.6	59	19.4	125	41.1	109	35.9	3.09	0.83

Table 14

Teacher Attitudes about Use of Benchmark Assessment Data and Professional Environment

Item	Strongly Disagree		Disagree		Agree		Strongly Agree		*M*	*SD*
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*		
Professional Environment										
Teachers in this school are continually learning and seeking new ideas.	10	3.3	31	10.1	171	55.7	95	30.9	3.14	0.72
Teachers are engaged in systematic analysis of student performance data.	10	3.4	65	21.8	179	60.1	44	14.8	2.86	0.70
Teachers in this school approach their work with inquiry and reflection.	10	3.3	40	13.1	172	56.2	84	27.5	3.08	0.73
Assessment of student performance leads to changes in the curriculum.	24	8.0	108	36.1	123	41.1	44	14.7	2.63	0.83
Teachers in this school regularly examine school performance on assessments.	8	2.7	50	16.7	177	59.2	64	21.4	2.99	0.70

Factors Influencing Teachers' Data Use

In addition to analyzing the descriptive patterns of teachers' survey responses, data analysis procedures also involved exploring and identifying relationships among the data that could further inform specific practices to support teachers' instructional use of benchmark test data. Prior to conducting correlational and regression analyses, the survey data were reduced into composite variables using theoretical approaches and empirical scales based on principle components factor analytic (PCA) procedures which employed a Varimax or orthogonal rotation technique. The PCA results showed that several underlying scales were present in the survey data. These scales reflected constructs related to organizational or conditional variables and

outcome variables related to teachers' specific use of data. See Appendix B for a summary of the PCA results, the items comprising each scale, and the factor loadings.

Condition variables. This section describes each of the variables that were created to capture key constructs or necessary conditions that have been shown in the literature to influence teachers' use of data to inform instructional decisions.

Alignment. The alignment composite variable is comprised of five survey items related to the alignment of benchmark assessments with state and district content standards, the state assessment, school district pacing guides, the content of instruction, and the appropriate level of difficultly for students. An example item includes, "The district benchmark assessments are well-aligned with state and division standards." Response options on these items ranged from "strongly disagree" to "strongly agree." Mean values on this variable range from 1 to 4 with values closer to 1.0 indicating strong disagreement and values closer to 4.0 indicating strong agreement. The internal consistency of this variable was sufficiently high with Cronbach's α =.90 (n = 300). As shown in Table 15, the mean on this variable was 2.83, suggesting that teachers were in general agreement that the benchmark assessments were aligned with the content of the state and district content standards as well as the content of their classroom instruction.

District policy. This composite variable is comprised of four survey items related to the school district's implementation of benchmark assessments. Example items include, "The district sets clear, consistent goals for schools to use data for school improvement," and "The district's data use policies help us address student needs at our school." Response options on these items ranged from "strongly disagree" to "strongly agree." Mean values on this variable range from 1 to 4 with values closer to 1.0 indicating strong disagreement and values closer to

4.0 indicating strong agreement. The internal consistency of this variable was sufficiently high with Cronbach's α =.86 (*n* = 267), and the mean was 2.44 indicating that teachers' views on the clarity and effectiveness of district policies related to benchmark assessments were mixed (see Table 15).

School environment. This composite variable is comprised of five survey items intended to measure the school climate related to learning, reflective practice and use of data. Example items include, "Teachers in this school are continually learning and seeking new ideas," "Teachers are engaged in systematic analysis of student performance data," and "Teachers in this school approach their work with inquiry and reflection." Response options on these items ranged from "strongly disagree" to "strongly agree." Mean values on this variable range from 1 to 4 with values closer to 1.0 indicating strong disagreement and values closer to 4.0 indicating strong agreement. The internal consistency of this variable was sufficiently high with Cronbach's α =.86 (*n* = 283). The mean for the school environment variable was 2.94, indicating that generally teachers were in agreement that the culture of the school facilitated and supported the use of benchmark assessment data (see Table 15).

Time spent analyzing data. This variable was derived to provide a composite measure of the time teachers spend analyzing and reviewing benchmark assessment data. The variable is comprised of five items related to the time teachers reported analyzing test results independently, with other teachers, the principal or assistant principal, students and parents. Response options on this variable included five different time increments ranging from 0, <1 hour; 1-2 hours; 2-3 hours; to more than 3 hours. Mean values range from 1 to 5, with values approaching 5.0 indicating larger amounts of time spent analyzing and reviewing benchmark data and values closer to 1.0 indicating smaller amounts of time. The internal consistency of this variable was

reasonably high with Cronbach's α =.72 (*n* = 358). As suggested by the mean value (2.84), the majority of teachers spent between 1-2 hours reviewing and analyzing the results of the most recently administered benchmark assessment (see Table 15).

Frequency of review and analysis. This conditional variable was intended to provide a composite measure of how often teachers analyzed and reviewed benchmark assessment data. The variable is comprised of eight survey items that measured the frequency of review with different groups or individuals including: the department or grade-level chair, the grade-level lead teacher, other teachers, instructional coaches, school administrators, central office staff, parents/guardians and students. The response options included a frequency scale ranging from Never; 1-2 times a quarter; 1-2 times a month; and 1-2 times a week. Mean values on this variable range from 1 to 4, with values closer to 4.0 indicating greater frequency of analysis and review and values closer to 1.0 suggesting less frequent analysis and review. The internal consistency of this variable was sufficiently high with Cronbach's α =.82 (*n* = 200). The mean value for this variable was 1.95 suggesting that teachers typically reviewed and analyzed results 1-2 times per month (see Table 15).

Teachers' interactions. This composite variable captures the range of teachers' interactions with others about benchmark assessment data. The variable is comprised of four items that asked teachers about the extent to which they engaged in different practices related to data. These practices included: meeting with grade-level teams or department teams to look at trends in the data (or analyze data); share ideas about using data to improve teaching with other teachers; share and discuss student work with other teachers; and discuss particular lessons that were unsuccessful. The response options included a four-point Likert-type scale ranging from "not at all" to "a major extent." Mean values closer to 1.0 indicate that teachers did not engage

in these types of interactions while values closer to 4.0 suggest that teachers engaged in a variety of interactions based on benchmark assessment data. The internal consistency of this variable was sufficiently high with Cronbach's α =.87 (n = 361). As shown in Table 15, the mean value for this scale was 2.84 indicating that on average teachers were likely to engage in several different types of interactions when conducting their review and analysis of the benchmark assessment data.

Teachers' use of benchmark data variables. Below is a description of several different variables that describe teachers' different uses of benchmark data for instruction. The variables measure constructs related to the instructional changes teachers make and the different types of instructional strategies they use on the basis of benchmark assessment data as well as the range of test score information that they incorporate into their decision making.

Instructional adjustments. This scale is comprised of 13 items that capture different types of instructional practices and changes made to instruction based on student benchmark assessment performance. The practices are generally related to adjusting instruction, changing curricular materials and student groupings. For example, teachers were asked how much influence division benchmark assessments had on the following: adjusting goals for student learning, determining a student's grouping for instruction, instructional strategies, adjusting pacing in areas where students encountered problems, and changing the sequence of instruction. Response options included a four-point Likert-type scale ranging from "no influence" to "a major influence." Mean values approaching 4.0 suggest that the benchmark assessments were a strong driver of instructional change compared to mean values closer to 1.0 which suggest that the division benchmark assessment was not influential when making instructional adjustments. The internal consistency of this variable was sufficiently high with Cronbach's α =.90 (n = 303).

The scale mean was 2.31 indicating that for most teachers, the benchmark assessments had some influence on their instructional decisions and adjustments (see Table 15).

Authentic instructional strategies.   This scale variable is comprised of eight different instructional approaches considered to encourage authentic or "real-world" problem-based learning. Teachers were asked to indicate how their "review of benchmark assessment results led [them] to decrease or increase the use of each of the following in their classroom instruction." The instructional approaches loading on this scale included: inquiry/investigation, problem solving activities, project-based assessments, use of student response journals, collaborative/team teaching, peer or cross-age tutoring, use of portfolios, and cooperative learning/group work. Response options included a five point Likert-type scale ranging from "large decrease" to a "large increase," with "no change" at the mid-point of the response scale. Mean values approaching 5.0 suggest that teachers' analysis of benchmark assessment results contributed to increased use of these instructional approaches while mean values closer to 1.0 suggest teachers' analyses led to decreased use of these strategies. The internal consistency of this variable was sufficiently high with Cronbach's α =.82 (n = 310). The mean on this scale was 3.56, indicating that on average teachers increased their use of authentic instructional strategies in response to the benchmark assessment results (see Table 15).

Traditional instructional strategies.   Similar to the authentic instructional approaches scale, this scale variable captures the influence of teachers' analysis of benchmark assessment data on the use of more traditional instructional approaches such as using lectures, worksheets, and text-book based assignments to promote student learning. Response options included a five point Likert-type scale ranging from "large decrease" to a "large increase," with "no change" at the mid-point of the response scale. Mean values approaching 5.0 suggest that teachers' analysis

of benchmark assessment results contributed to increased use of traditional strategies while mean values closer to 1.0 suggest teachers' analyses led to decreased use of these strategies. The internal consistency of this variable was modest with Cronbach's α =.61 ($n = 320$), the small number of items (n =3) comprising the scale contributed to the lower than desirable alpha value. The scale mean was 3.01, indicating that on average teachers did not modify the extent to which they used traditional instructional approaches as a result of student performance on the benchmark assessments (see Table 15).

Use of scores. This scale-level variable is comprised of five survey items that measured the use of specific types of score reporting of benchmark assessment data. Teachers were asked about the extent to which they used of the following types of benchmark assessment results. These scores included: results for subgroups of students, scale scores or other scores that show how close students are to performance levels, results for each grade level, results for specific reporting categories, percent of students scoring at or above the proficient level. Response options included a four-point Likert-type scale ranging from "did not use in this way" to "extensive use." Mean values closer to 4.0 suggest that teachers used a variety of different types of benchmark assessment scores in their analysis and review of the data, while values closer to 1.0 suggest a more limited approach to the range of data included in analysis. The internal consistency of this variable was reasonable with Cronbach's α =.79 ($n = 283$). The mean value for this scale was 2.53 indicating that teachers typically used multiple types of results and score reporting in their analysis of benchmark assessment results (see Table 15).

Relationships among conditions and teachers' use of benchmark assessment data. Bivariate correlations were computed for the six different condition composite variables and the four scale-level variables of different types of instructional uses of benchmark assessment data.

The results are shown in Table 15. As shown, significant positive correlations at $p < .01$ were

found for district policy ($r =.336$), school environment ($r =.218$), frequency of analysis and

review ($r =.425$), teachers' interactions ($r =.381$), time spent analyzing results ($r =.486$) and

instructional adjustments. These data suggest that increased efforts to analyze and review

benchmark assessment data is associated with increased instructional changes as well as

increased use of authentic, problem-based approaches to learning. Similarly, it follows that

increased frequency and time spent analyzing benchmark assessment results is associated with

the use of a wider range of data as suggested by the positive correlations of frequency of analysis

and review ($r = .400$), time spent analyzing ($r = .398$), and the use of scores.   Also noteworthy,

is the lack of statistically significant correlations found among the six conditions and the

traditional instructional strategies scale variable.

Table 15

Bivariate Correlations among Conditions and Teachers' Use of Results

Condition	Instructional Adjustments	Authentic Instructional Strategies	Use of Scores	Traditional Instructional Strategies	M	SD
Alignment	.229*	.101	.266*	.007	2.83	0.65
District Policy	.336**	.087	.373**	-.022	2.44	0.66
School Environment	.218**	.084	.189*	.088	2.94	0.59
Frequency of Review and Analysis	.425**	.249**	.400**	.036	1.95	0.53
Teachers' Interactions	.381**	.113*	.398**	-.039	2.66	0.74
Time Spent Analyzing	.486**	.186**	.398**	.028	2.84	0.75
M	2.31	3.56	2.53	3.01		
SD	0.67	0.62	0.76	0.64		

Note. *correlations significant at $p < .05$; **correlations significant at $p < .01$.

In order to examine the conditions most predictive of teachers' use of benchmark

assessment data, stepwise multiple linear regressions were conducted. The stepwise regression

technique allows for the prediction of one outcome or dependent variable from several different

independent or predictor variables, with the final, most predictive model shown in the last step.

Stepwise regression procedures were conducted to determine which combination of the six

different organizational conditions predicted the degree to which teachers made instructional

adjustments and used benchmark assessment scores. For instructional adjustments, the results

indicate that the model including the following four conditions - time spent analyzing data,

district policy, frequency of reviewing benchmark data, and teacher interactions - accounted for

31.2% of the variance; $F = 30.857$, $p < .001$. Time spent analyzing benchmark data which

accounted for the largest amount of variance ($r^2 = .239$, $\beta = .310$); $F = 62.36$, $p < .001$. Table 16

provides the regression results.

Table 16

Stepwise Regression of Conditions Predicting Instructional Adjustments

Model	Conditions	R	R^2	Beta	Sig.
1	Frequency of Review	.430	.185	.430	.000
2	Teachers Interactions	.484	.234	.253	.000
3	Time Spent Analyzing	.541	.293	.310	.000
4	District Policy	.559	.312	.153	.006

Similarly, to examine which conditions were most predictive of teachers' use of

benchmark assessment scores, stepwise multiple linear regression techniques were conducted.

This analysis examined the degree to which the six organizational conditions predicted teachers'

use of specific benchmark assessment scores. For use of scores, the final model includes four of

the six conditions – frequency of review, teacher interactions, time spent analyzing data, and district policy (see Table 17).

Table 17

Stepwise Regression of Conditions Predicting Use of Benchmark Scores

Model	Conditions	R	R^2	Beta	Sig.
1	Frequency of Review	.412	.169	.412	.000
2	Teacher Interactions	.481	.232	.282	.000
3	Time Spent Analyzing	.500	.250	.174	.014
4	District Policy	.535	.286	.209	.000

The model accounted for 28.6% of the variance; $F = 25.207$, $p < .001$). Although the model including the four conditions explained the most variance, a large percentage remains unexplained, suggesting that other factors may account for the extent teachers use the different benchmark assessment score information.

Discussion

The results of the survey demonstrate that teachers are using benchmark assessment data to inform decisions about students and their own instructional practice. The results indicate they are most likely to change their teaching method and add or change the skills emphasized in their instruction on the basis of benchmark assessment results. Teachers reported adjusting pacing in areas where students encountered problems, suggesting a response for future instruction rather than an immediate one to address student learning needs. These results indicate that teachers are making more procedural or surface level adjustments as a result of benchmark assessment data. However, when asked about how much they increased or decreased certain instructional

approaches, their responses suggested that the benchmark assessment results were increasing their use of more real-world authentic learning experiences. Teachers reported in large percentages that they increased time spent on problem-solving activities (58%), cooperative learning (49%), strategies that involve inquiry and investigation (47%), peer tutoring (31%) and collaborating or team teaching (29%). Very few teachers, roughly 8%, reported increase time spent on worksheets, text-book based assignments or lectures and on average, teachers reported making no changes to the use of traditional instructional approaches. These findings are suggestive of instructional changes based on students' conceptual misunderstandings rather than their more rote or procedural responses described in our earlier focus group sessions or in other survey-based research (Goertz et al., 2009; Oláh et al., 2010; Shepard et al., 2011). These apparent inconsistencies may indicate that teachers view the need to address learning needs identified by benchmark assessment data as separate from their regular day-to-day instruction. Teachers may be spiraling the curriculum where they are using data to identify conceptual problems and are addressing deficit skills or knowledge when teaching new curriculum.

Teachers' instructional use of benchmark assessment data is driven by their approach to analysis. Teachers were most likely to report using results according to different content standards and reporting categories as well as the percentage of students at different levels of proficiency in their analysis. They were less likely to analyze results of different subgroups of students according to Annual Yearly Progress (AYP) categories for example. These results suggest that teachers are engaging in more cursory forms of data analysis rather than finer-grained analysis were student characteristics are intersected with proficiency levels or reporting categories. This finding is consistent with other research conducted in this same locality. Hoover and Abrams (2011) found from their district-wide survey of 650 teachers that they most

frequently relied on measures of central tendency (e.g., average, mode, and median) and variability (e.g., standard deviation) in their analysis of assessment data. Teachers were also more likely to report analyzing assessment data by content standards rather than by AYP subgroups. About a third of teachers reported never analyzing data by subgroup. Of those who did, this type of analysis occurred most often with benchmark assessment data. Teachers' reliance on summary level results suggests that the information obtained from the analysis of interim assessment data is limited and as such provides limited direction for instructional responses. Bernhardt (2000, 2004) describes the power of data analysis comes from the intersection of data sets where the points of intersection often reveal information most useful for instruction and learning. The survey results suggest that there is a continued need to develop not only teachers', but also building administrators', expertise in data analysis and interpretation as well as provide the time necessary to engage in thoughtful analysis and discussion.

Another goal of this study was to explore if teachers were using benchmark assessment data formatively. When considered in light of the 2008 CCSSO definition of formative assessment, the results suggest that benchmark assessment data are being used to make instructional adjustments with the intent of improving student learning outcomes. At a very basic level, we can conclude that teachers are engaging in formative assessment based on their reported use of benchmark assessment data. However, when considered according to the theory of action of formative assessment put forth by Black and Wiliam (2009), the extent to which teachers' reported use of interim assessment data is formative is less conclusive. According to their five essential components of formative assessment, we can argue that teachers are using interim assessment data to clarify learning intentions and that these intentions are shared with students either directly or indirectly through re-teaching and remediation efforts. Teachers also

46

seem to engage students as resources for one another through grouping strategies as well as reports of implementing peer tutoring in their classrooms. What is less evident are the formative assessment practices associated with the remaining three theoretical components, especially the degree to which students are receiving feedback on the basis of the test results and if a formal review of results provides sufficient direction to move learners forward. To draw firm conclusions about the formative nature of teachers' use of benchmark assessment results, as defined by Black and Wiliam's theory of action, more questions need to be asked and further study undertaken. This theory of action encourages greater depth in research on interim assessments and formative assessment. Such work would also address some of the limitations identified in the literature related to the perhaps tenuous connections between instructional responses and student misunderstandings (Goertz et al., 2009; Oláh et al., 2010; Shepard et al., 2011).

Based on the results of the two phases of our empirical work on teachers' use of benchmark assessment data, we offer several recommendations for effective use of data to inform teaching that is intended to address students' misunderstandings of curriculum. The recommendations reflect general principles of high quality assessment established by the *Standards for Educational and Psychological Testing* (American Educational Research Association, the American Psychological Association, & National Council on Measurement in Education, 1999). In addition to sound measurement principles and practices, our recommendations also reflect what we heard from teachers and administrators about policies that could support and build their capacity to effectively use data in formative ways. These include for example, providing teachers with the test questions and answer options along with the results, allocating time during the school day for teachers to analyze and discuss results and the relevant

applications to instructional strategies, organizing professional learning communities or teams of teachers to meet for the purpose of discussing and analyzing benchmark assessment data. It is also clear that school divisions need to address the tension between pacing pressures and using benchmark assessment results to re-teach or remediate in ways other than adding time to the school day. Enhancing teachers' expertise in data analysis and interpretation through increased opportunities for professional development or through data coaches may provide for greater capacity and benefit broader school improvement efforts.

Recommendations

1. Clarify the purpose of the benchmark assessments with all stakeholders to communicate a singular purpose to use results to make instructional adjustments.

2. Establish alignment evidence with the content and cognitive level of the state curriculum standards and with district pacing guides.

3. Establish district and school environments that support data-driven decision making.

4. Use high quality test items that provide for valid interpretations and inferences about student learning.

5. Provide structured time for teams of teachers to review and analyze results during the school day.

6. Distribute test questions along with results which should show the numbers of students selecting each response option.

7. Provide adequate professional development to support teachers' use of results.

8. Evaluate use of test results to determine if evidence exists that teachers are using results to modify instruction and that students' learning is improving. Verify these results with other sources of achievement evidence.

9. Standardize policies and test administration procedures for all schools within a district.

10. Document costs – How much instructional time is being replaced by testing, test preparation, review and analysis of results? How much does the benchmark assessment program cost in terms of software and personnel resources?

Conclusion

The literature on data-driven decision making and formative assessment provides for a strong foundation on which districts, schools, and teachers can develop models of inquiry and reflective practice that are most closely aligned with formative assessment. There is compelling evidence of the potential for benchmark assessment data to have a profound impact on instruction and in turn student learning. We know that teachers are using data, most often to identify common student misunderstandings and relate these misconceptions to content standards. To address these learning deficits teachers often re-teach in large or small groups depending on the degree of student misunderstandings and they will often provide remediation most commonly before or after school to address highly individualized student needs. We also know that gaps persist in how teachers are using benchmark assessment data to make instructional adjustments that directly align with student misconceptions. Increased attention on the relationship between student misunderstandings as identified by analysis of benchmark assessment data, and the nature of instructional responses as well as the capacity of benchmark assessments to provide information about conceptual knowledge are needed and should inform future research on teachers' use of benchmark assessment data.

<div align="center">**References**</div>

Abrams, L., Wetzel, A., & McMillan, J. (2010). *Formative assessment practices with benchmark*

    testing: Phase 1. Richmond, Virginia: Metropolitan Educational Research Consortium,

    Virginia Commonwealth University.

Bennett, R. (2011). Formative assessment: A critical review. *Assessment in Education:*

    Principles, Policy & Practice, 18(1), 5-25.

Bernhardt, V. L. (2000). Intersections: New routes open when one type of data crosses another.

    Journal of Staff Development, 21(1), 33–36.

Bernhardt, V. L. (2004). Continuous improvement: It takes more than test scores. *ACSA*

    Leadership, 16–19. Retrieved from http://eff.csuchico.edu/downloads/testscores.pdf

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education,*

    5(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing a theory of formative assessment. *Educational*

    Assessment, Evaluation & Accountability, 21, 5-31. DOI: 10.1007/s11092-008-9068-5.

Blanc, S., Christman, J., Liu, R., Mitchell, C., Travers, E., & Bulkley, K. (2010). Learning to

    learn from data: Benchmarks and instructional communities. *Peabody Journal of*

    Education, 85(2), 205-225.

Brunner, C., Fasca, C., Heinze, J., Honey, M., Light, D., & Mandinach, E. (2005). Linking data

    and learning: The grow network study. *Journal of Education for Students Placed at*

    Risk, 10(30), 214-267.

Burch, P. (2010). The bigger picture: Institutional perspectives on interim assessment

    technologies. *Peabody Journal of Education, 85*(2), 147-162.

Christman, J., Neild, R., Bulkley, K., Blanc, S., Lui, R., Mitchell, C., & Travers, E. (2009).

    Making the most of interim assessment data: Lessons from Philadelphia. Philadelphia:

    Research for Action.

Datnow, A., Park, V., & Wohlstetter, P. (2007*). Achieving with data*. Los Angeles, California:

    Center on Educational Governance, University of Southern California.

Duschl, R., & Gitomer, D. (1997). Strategies and challenges to change the focus of assessment

    and instruction in science classrooms. *Educational Assessment, 4*(1), 37-73.

Dunn, K., & Mulvenon, S. (2009). A critical review of research on formative assessment: The

    limited scientific evidence of the impact of formative assessment in education. *Practical*

    Assessment, Research & Evaluation, 14(7). Available online:

    http://pareonline.net/getvn.asp?v=14&n=7

Fennema, E., Franke, M., Carpenter, T., & Carey, D. (1993). Using children's mathematical

    knowledge in instruction. *American Educational Research Journal, 30*(3), 555-583.

Halverson, R., Grigg, J., Pritchett, R., & Thomas, C. (2005). The new instructional leadership:

    Creating data-driven instructional systems in school. *Journal of School Leadership, 17*,

    159-194.

Henderson, S., Petrosino, A. ,Guckenburg, S., & Hamilton, S. (2008). *A second follow-up year*

    for "Measuring How Benchmark Assessments Affect Student Achievement". (REL

Technical Brief, REL 2008-No. 002). Regional Educational Laboratory Northeast &
Islands.

Hoover, N., & Abrams, L. (2011, April). *Teachers' instructional use of student assessment data*.
Paper presented at the meeting of the American Educational Research Association, New
Orleans, LA.

Kerr, K., Marsh, J., Ikemoto, G., Darilek, H., & Barney, J. (2006). Strategies to promote data
use of instructional improvement: Actions, outcomes and lessons from three urban
districts. *American Journal of Education*, *112*, 496-520.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research.
Educational Measurement: Issues and Practices, 30(4), 28-37.

Lachat, M., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal
of Education for Students Placed at Risk, 10*(3), 333-349.

Marsh, J., Pane, J., & Hamilton, L. (2006). *Making sense of data-driven decision making in
education*: *Evidence from recent RAND research*. Washington, DC: RAND
Corporation.

Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009*). Implementing data-informed decision
making in schools – teacher access, supports and use*. Washington, DC: U.S.
Department of Education, Office of Planning, Evaluation, and Policy Development.

Murnane, R., Sharkey, N., & Boudett, K. (2005). Using student-assessment results to improve
instruction: Lessons from a workshop. *Journal of Education for Students Placed at Risk,
10*(3), 269-280.

Niemi, D., Wang, J., Wang, H, Vallone, J., & Griffin, N. (2007). *Recommendations for building a valid benchmark assessment system: Second report to the Jackson Public Schools*. (CRESST Report No. 724). Retrieved from the National Center for Research on Evaluation, Standards, & Student Testing website: http://www.cse.ucla.edu/products/reports/R724.pdf

Nelson, M., & Eddy, R. (2008). Evaluative thinking and action in the classroom. In T. Berry & R. Eddy (Eds.). *Consequence of No Child Left Behind for Educational Evaluation: New Directions for Evaluation, 177*, 37-46.

Oláh, L., Lawrence, N., & Riggan, M. (2010). Learning to learn from benchmark assessment data: How teachers analyze results. *Peabody Journal of Education, 85*(2), 226-245.

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. Washington, DC: The Aspen Institute.

Riggan, M., & Oláh, L. (2011). Locating interim assessments within teachers' assessment practice. *Educational Assessment, 16*(1), 1-14.

Shepard, L., Davidson, K., & Bowman, R. (2011). *How middle school mathematics teachers use interim and benchmark assessment data* (CRESST Report No. 807). Retrieved from the National Center for Research on Evaluation, Standards, & Student Testing website: http://www.cse.ucla.edu/products/reports/R807.pdf

Symonds, K. (2004). *After the test: Closing the achievement gaps with data*. San Francisco, CA: Learning Point Association and Bay Area School Reform Collaborative.

Supovitz, J., & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement.* Philadelphia: Consortium for Policy Research in Education.

Trimble, S., Gay, A., & Matthews, J. (2005). Using test score data to focus instruction. *Middle School Journal, 36*(4), 26-32.

Vogel, L., Rau, W., Baker, P., & Ashby, D. (2006). Bringing assessment literacy to the local school: A decade of reform initiatives in Illinois. *Journal of Education for Students Placed at Risk, 11*(1), 39-55.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement, 19*(3), 239-259.

Yeh, S. (2006). High-stakes testing: Can rapid assessment reduce the pressure? *Teachers College Record, 108*(4), 621-661.

MERC Benchmark Survey

SECTION 1: BACKGROUND

What grade(s) do you currently teach (please check all that apply):
- Fourth
- Fifth
- Sixth
- Seventh
- Eighth
- Other (please specify) _____

What subjects(s) do you primarily teach? (Please check all that apply):
- Reading/English Language Arts
- Mathematics
- Science
- Social Studies
- Special Education (in a self-contained classroom, resource room, or inclusion classroom)
- English as a Second Language

How many years of teaching experience do you have in each of the following settings? Include any full-time teaching assignments, part-time teaching assignments, and long-term substitute assignments.

Special Instruction: For each row, enter the number of years in whole numbers only, and count the current school year as one year.

Total number of years teaching (including this year)
(_____)

Number of years teaching in this school (including this year)
(_____)

What is your gender?
- Male
- Female

Are you of Hispanic or Latino origin?
- Yes
- No

What is your race?
- White
- Black or African American
- Asian
- Native Hawaiian or other Pacific Islander
- American Indian or Alaska Native

Have you earned any of the following degrees, diplomas, or certificates?
(Please select yes or no, select your major, and the year you completed the degree)

Bachelor's Degree - Earned
- Yes
- No

Bachelor's Degree - Major
(_____)

Bachelor's Degree - Year
(_____)

Master's Degree - Earned
- Yes
- No

Master's Degree - Major
(_____)

Master's Degree - Year
(_____)

Educational specialist or professional diploma (at least one year beyond master's level) - Earned
- Yes
- No

Educational specialist or professional diploma (at least one year beyond master's level) - Major
(_____)

Educational specialist or professional diploma (at least one year beyond master's level) - Year
(_____)

Certificate of advanced graduate studies - Earned
- Yes
- No

Certificate of advanced graduate studies - Major
(_____)

Certificate of advanced graduate studies - Year
(_____)

Doctorate or professional degree (Ph.D., Ed.D., M.D., L.L.B., J.D., D.D.S) - Earned
- Yes
- No

Doctorate or professional degree (Ph.D., Ed.D., M.D., L.L.B., J.D., D.D.S) - Major
(_____)

Doctorate or professional degree (Ph.D., Ed.D., M.D., L.L.B., J.D., D.D.S) - Year
(_____)

Number of years teaching in this division (including this year)
(_____)

In which school division do you work?
- Chesterfield County Public School
- Colonial Heights City Schools
- Hanover County Public Schools
- Hopewell City Public Schools
- Powhatan County Public Schools
- Richmond City Public Schools
- Goochland County Public Schools

SECTION 2: DIVISION/SCHOOL POLICIES AND PROCEDURES

Which of these types of assessments are you required by your division or this school to administer on a periodic basis (e.g., every 4-9 weeks) to monitor your student's progress? *Please check all that apply.*
- Assessments created by myself or others in my school
- Assessments from the curriculum program (e.g. curriculum-based unit assessments)
- Benchmark assessments developed by or for our division
- Other commercial assessments
- Other [_____]

Please indicate how much you agree or disagree with the following statements.

The principal at my school:
Scale for the items in this section:
- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

1. Encourages teachers to make decisions based on data.
2. Takes primary responsibility for presenting and interpreting benchmark assessment results for teachers.
3. Places too much emphasis on benchmark assessment results.
4. Commits resources to help teachers interpret and use benchmark assessment data.

Since the last benchmark test administration, about how often did your school have scheduled meeting time to:
Please select one option from the following drop-down choices:
- About once a week
- 1 - 2 times per month
- 1 - 2 times per quarter
- My school does not provide time for this
- N/A

1. Review benchmark assessment data - As Grade-Level Teams or as Departments
2. Review benchmark assessment data - As a Whole Staff
3. Review other types of student data (e.g. state assessment scores, student work, attendance, etc.) - As Grade-Level Teams or as Departments
4. Review other types of student data (e.g. state assessment scores, student work, attendance, etc.) - As a Whole Staff

5. Discuss student achievement by subgroup (e.g. students with disabilities, ELL/LEP, gender, race/ethnicity) - As Grade-Level Teams or as Departments
6. Discuss student achievement by subgroup (e.g. students with disabilities, ELL/LEP, gender, race/ethnicity) - As a Whole Staff
7. Discuss individual student achievement - As Grade-Level Teams or as departments
8. Discuss individual student achievement - As a Whole Staff
9. Discuss and share instructional strategies - As Grade-Level Teams or as departments
10. Discuss and share instructional strategies - As a Whole Staff
11. Meet with an instructional coach - As Grade-Level Teams or as departments
12. Meet with an instructional coach - As a Whole Staff
13. Meet with a data coach - As Grade-Level Teams or as departments
14. Meet with a data coach - As a Whole Staff

SECTION 3: ACCESSING BENCHMARK DATA

How do you primarily access benchmark assessment results?
Please select one:
- Immediately after the student completes the assessment (as in computer-adaptive assessment)
- Online, through a web-based system or database
- Electronic reports provided by the school
- I access results electronically and then print out hard copies
- Hard-copy reports provided by the school
- I never access benchmark assessment results

Are the assessment questions provided with the results?
- Yes
- No

How long after the administration of the most recent benchmark assessments were results made available to you?
- Immediately (within 24 hours)
- Within 2 to 3 days
- Within 1 week
- Within 2 weeks
- It takes longer than 2 weeks
- N/A - no results were made available to me

SECTION 4: ANALYZING RESULTS

How frequently do you review student benchmark assessment data with the following people?
- Never
- 1 or 2 times a Quarter
- 1 or 2 times a Month
- 1 or 2 times a Week
- N/A

1. Department chair/grade-level chair
2. Grade level lead teacher
3. Other classroom teachers in my grade level or subject area
4. Instructional coaches
5. School administrators
6. Division central office staff
7. Parents/guardians
8. Students

This question concerns how teachers interact with each other in your school as related to benchmark assessment data. Please indicate the extent to which you do each of the following.
- Not At All
- Slight Extent
- Moderate Extent
- Major Extent

1. Meet with grade-level teams or department teams to look at trends in the data (or analyze data)
2. Share ideas about using data to improve teaching with other teachers
3. Share and discuss student work with other teachers
4. Discuss particular lessons that were not very successful

Please select one option from the following drop-down choices to answer the question:

Since the most recent benchmark assessment, approximately how many hours did you engage in analyzing and/or reviewing benchmark assessment data in the following ways?
- 0
- <1 hour
- 1-2 hours
- 2-3 hours
- More than 3 hours

1. Independently
2. With teachers in my grade level
3. With teachers in other grade levels
4. With my principal
5. With my assistant principal
6. With a content-area coach (e.g. math or reading coach)
7. With a data coach
8. With students
9. With parents
10. Other; please specify

How much have you used the most recent benchmark assessment results to-
Scale for the following items is:

- Did Not Use In This Way
- Minimal Use
- Moderate Use
- Extensive Use

1. Identify individual students who need remedial assistance
2. Diagnose learning problems
3. Tailor instruction to individual students' needs
4. Identify and correct gaps in the curriculum for all students
5. Recommend tutoring or other educational services for students
6. Identify areas where I need to strengthen my content knowledge or teaching skills
7. Assign or reassign students to classes or groups
8. Determine instructional materials to use with my class(es)
9. Develop or revise Individualized Education Programs (IEPs)

How useful were these types of review/analysis of benchmark assessment data for your teaching?

- Not at all Useful
- Not Very Useful
- Somewhat Useful
- Very Useful

1. Independent
2. With teachers in my grade level
3. With my principal
4. With my assistant principal
5. With content-area coach (e.g. math or reading coach)
6. With a data coach
7. With students
8. With parents

SECTION 5: INSTRUCTIONAL USE OF RESULTS

To what extent do the following factors influence your classroom instruction?

- No Influence
- Minor Influence
- Moderate Influence
- Major Influence
- N/A

1. The state's curriculum framework or content standards
2. Your division's curriculum framework, standards, or guidelines
3. Division policies and initiatives
4. Benchmark assessment data
5. Assessments that you develop
6. Curriculum-based unit assessments
7. End-of-year state assessment scores
8. Division pacing guides
9. Your own classroom observations

Consider the reports you receive or generate for students' benchmark assessment results.

How helpful are the following:

- Somewhat Helpful
- Not Very Helpful
- Very Helpful
- Not Helpful
- Not Reported

1. The number of students that correctly answer each test item
2. The percentage of students that correctly answer each test item
3. The assessment norms for students that correctly answer each test item
4. The results presented according to different reporting categories
5. The results that include test items keyed to SOLs
6. Assessment results for individual student results

On average, how often do you use benchmark assessment data to:

- Never
- 1 - 2 times per year
- 3 - 4 times per year
- About monthly
- Weekly
- Daily

1. Inform students of their progress
2. Involve students in interpreting their own benchmark assessment results
3. Involve students in creating new strategies for learning based on benchmark assessment data
4. Inform parents of student progress

To what extent do you use the following types of benchmark assessment results?

- Did Not Use In This Way
- Minimal Use
- Moderate Use
- Extensive Use
- Not Made Available In This Way

1. Percent of students scoring at or above the proficient level
2. Scale scores or other scores that show how close students are to performance levels
3. Results for subgroups of students (e.g. students with disabilities, ELL/LEP, gender, race/ethnicity)
4. Results for each grade level
5. Results for your class(es)
6. Results on specific reporting categories
7. Item-by-item results

To what extent did you do the following to address the needs of students as a direct result of students' benchmark assessment scores?

- Not At All
- Minor Extent
- Moderate Extent
- Major Extent

1. Reviewed key concepts for the entire class
2. Used same-level achievement groupings
3. Used mixed-level achievement groupings
4. Used individualized instruction during class to address the needs of struggling students
5. Provided individual assistance outside of class to address the needs of struggling students
6. Changed the sequence of instruction
7. Added, deleted, or changed skills taught
8. Changed teaching method (e.g. lecture, cooperative learning, student inquiry)

How much influence do division benchmark assessment results have on the following?

- No Influence
- Minor Influence
- Moderate Influence
- Major Influence

1. Determining a student's grouping for instruction
2. Adjusting goals for student learning
3. Adjusting pacing in areas where students encountered problems
4. Adjusting use of textbooks and instructional materials
5. Identifying students to be retained at the same grade level
6. Identifying students for tutoring or other supplemental instruction
7. Identifying professional development needs
8. Evaluating teachers

Based on the most recent benchmark assessment results, how much have you changed the following aspects of your teaching?

- No Change
- Minor Change
- Moderate Change
- Major Change

1. The curriculum content I teach
2. My expectations for student performance
3. The instructional strategies I employ
4. The types of mix of assessments I use to evaluate students

---

To what extent has your review of benchmark assessment results led you to decrease or increase the use of each of the following in your classroom instruction?

- Large Decrease
- Decrease
- No Change
- Increase
- Large Increase
- N/A

1. Writing assignments
2. Textbook based assignments
3. Inquiry/Investigation
4. Problem-solving activities
5. Worksheets
6. Project-based assignments
7. Use of student response journals
8. Use of portfolios
9. Lecturing
10. Cooperative learning/group work
11. Computers/educational software
12. Peer or cross-age tutoring
13. Collaborative/team teaching

To what extent have the following factors hindered your ability to use data to make instructional decisions based on benchmark assessment data?

- Not At All
- Minor Extent
- Moderate Extent
- Major Extent

1. Lack of time to study and think about available data
2. Lack of time to collaborate with others in analyzing and interpreting data
3. Not enough professional development
4. Personal discomfort with data analysis
5. Lack of technology (e.g. access to computer with reliable internet connection)
6. Insufficient amount of data
7. Data provided too late for use
8. Curriculum pacing pressures
9. Division pacing guides do not allow me to re-teach based on results of benchmark assessment
10. Other

Since the last benchmark assessment, what percentage of instructional time did you spend on the following activities?

- Less than 1%
- 2 - 5 %
- 6 - 10%
- 11 - 15%
- 16 - 20%
- 21 - 25%
- More than 25%

1. Teaching specific test-taking strategies or skills
2. Administering practice tests or quizzes that mirror the quarterly benchmark test
3. Administering the division benchmark test
4. Reviewing benchmark test results with students
5. Other, please specify: (_____)

SECTION 6: ATTITUDES TOWARD BENCHMARK TESTS

Thinking about the most recent division benchmark assessments administered at your school, how much do you agree or disagree with the following statements?

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree
- Don't Know

The division benchmark assessments are:

1. Well-aligned with state and division standards
2. Well-aligned with the state assessment
3. Well-aligned with the pacing guides
4. Well-aligned with what I teach in the classroom
5. Appropriately challenging for my students
6. Of little use to me in my instruction

How much do you agree or disagree with the following statements about your division's priorities about benchmark assessment data?

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree
- Don't Know

1. The division sets clear, consistent goals for schools to use data for school improvement.
2. Division staff provide information and expertise that support the data use efforts at my school.
3. The division's data use policies help us address student needs at our school.
4. The division has designated adequate resources (e.g. time, staff, money) to facilitate teachers' use of data.

How much do you agree or disagree with the following statements about using division benchmark assessment data?

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree
- Don't Know
- N/A

1. Benchmark assessment results are reported to me in a timely manner.
2. Benchmark assessment data are easy to use.
3. The division provides benchmark assessment data to schools in easy-to-use formats.
4. It is easy to access benchmark assessment data directly in the division data system.
5. My school's internet connection enables teachers to assess the division benchmark assessment system online.
6. There are enough computers at my school to enable teachers to access the division benchmark assessment system online.
7. If I want to use benchmark assessment data in my teaching, I have to use my personal time to review the data.

Now consider the professional climate in your school. To what extent do you agree or disagree with each of the following statements?

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree
- Don't Know

1. Teachers in this school are continually learning and seeking new ideas.
2. Teachers are engaged in systematic analysis of student performance data.
3. Teachers in this school approach their work with inquiry and reflection.
4. Assessment of student performance leads to changes in the curriculum.
5. Teachers in this school regularly examine school performance on assessments.

Please rate your proficiency at the following activities:

- Not At All Proficient
- Barely or Slightly Proficient
- Moderately Proficient
- Very Proficient

1. Analyzing trends in student performance over time
2. Translating data into knowledge about student strengths and weaknesses
3. Incorporating benchmark assessment data into lesson planning
4. Using benchmark assessment data to adapt my teaching
5. Using student data by subgroup (students with disabilities, ELL/LEP, gender, race/ethnicity) to improve student performance

Appendix B

Summary of Principal Component Analysis

Table A1

Principal Components Analysis Results

Scale/Items on Scale	Loading	M	SD
Instructional Adjustments (α =.90).			
Adjusting goals for student learning	.758	2.40	0.95
Determining a student's grouping for instruction	.735	2.19	0.99
The types of mix of assessments I use to evaluate students	.710	2.24	0.94
The instructional strategies I employ	.698	2.40	0.91
Adjusting pacing in areas where students encountered problems	.689	2.56	0.96
Adjusting use of textbooks and instructional materials	.672	2.12	0.95
Changed teaching method (e.g. lecture, cooperative learning, student inquiry)	.600	2.57	.096
The curriculum content I teach	.650	1.89	0.91
Used same-level achievement groupings	.619	2.19	1.01
Changed the sequence of instruction	.617	2.30	1.01
Used mixed-level achievement groupings	.557	2.27	1.02
Added, deleted, or changed skills taught	.541	2.57	0.92

Scale/Items on Scale	Loading	M	SD
Authentic Instructional Strategies (α =.82)			
Inquiry/Investigation	.767	3.54	0.80
Problem-solving activities	.732	3.67	0.79
Project-based assignments	.697	3.31	0.82
Use of student response journals	.659	3.55	1.05
Collaborative/team teaching	.630	3.64	1.00
Peer or cross-age tutoring	.622	3.61	0.95
Use of portfolios	.616	3.61	1.21
Cooperative learning/group work	.602	3.55	0.74
Specific use of Scores			
Results for subgroups of students	.766	2.36	1.05
Scale scores or other scores that show how close students are to performance levels	.736	2.43	1.03
Results for each grade level	.724	2.23	1.07
Results on specific reporting categories	.698	2.77	1.03
Percent of students scoring at or above the proficient level	.662	2.85	.97
Traditional Teaching Methods			
Lecturing	.687	2.94	.73
Worksheets	.635	2.97	.69
Textbook based assignments	.563	3.14	1.09