Massachusetts Study on Assessment Practices in Districts: Phase 1 Summary

|  |  |
| --- | --- |
| **Date:** | March 14, 2015 |
| **To:** | Carrie Conaway, Massachusetts Department of Elementary and Secondary EducationKendra Winner, Massachusetts Department of Elementary and Secondary Education |
| **From:** | Julia MarchandCorrin PitluckSarah Cristobal |

In November 2014, the Massachusetts Department of Elementary and Secondary Education (ESE) partnered with American Institutes for Research (AIR) to conduct a study examining assessment preparation, practices, and supports in districts. The study, to be completed in spring 2015, will be carried out in two interrelated phases. The first phase builds on the findings from a recent ESE-administered statewide survey of district leaders on their assessment practices by gathering focused qualitative information from a representative sample of districts across the state about their practices, procedures, and strategies related to assessment. The study’s second phase will more closely examine assessments and assessment practices in a set of four districts that participated in Phase 1.

This memo summarizes findings from Phase 1 of the study, which took place from December 2014 through February 2015.

# Phase 1 Overview

The objectives of Phase 1 were to identify examples of promising or best practices related to assessment practices to share statewide and to identify potential districts for participation in the second phase of the study. To accomplish this, AIR researchers conducted telephone interviews with district- and school-level assessment leads in districts from across the state.

## Sample

The sample in this study was drawn with several purposes in mind: to obtain a representative and inclusive sample of districts from across the state and to ensure proportional numbers of elementary, middle, and high school administrators across participating districts are well-represented. Quota sampling—a nonprobabilistic method—was used. AIR researchers began with a matrix that described the target population’s characteristics: what proportion of all the districts is urban or nonurban, for example, and for the urbanicity classification, what proportions fall into various regions, accountability levels, district size, percentage of English language learners (ELLs), and school type (i.e., charter and traditional).

Once the matrix had been created and a relative proportion was assigned to each cell in the matrix, a quota sample of 40 school districts was established. The researchers then collected the list of schools within each sampled district. Another matrix was created to define the schools’ characteristics (e.g., grade levels served, student enrollment, and percentage of students with limited English proficiency). One school from each sampled district was selected after the schools were assigned a weight that is appropriate to their portion of the total population to ensure a reasonable representation. Figures 1 through 3 describe the sample of 40 districts drawn based on characteristics of interest to this study.

Figure 1. Districts by Accountability Level (*N* = 35)

Figure 2. Districts by Region (*N* = 35)

Figure 3. Districts by Urbanicity and Charter Status (*N* = 35)

Given the range of district-level data for specific continuous variables of interest, AIR researchers created unique categories from low to high that allow for the comparison of groups of districts. Table 1 specifies the overall and disaggregated ranges for each of the variables in question. Figures 4 and 5 present the data for each category.

Table 1. Disaggregated Variables

|  |  |  |  |
| --- | --- | --- | --- |
|  | Percentage of Special Education | Percentage of ELLs | Percentage of Proficient or Higher on MCAS 2014a |
| Overall range | 3–37% | 0–47% | 40–87% |
| Low | 0–10% | 0–10% | 40–55% |
| Medium  | 11–20% | 11–20% | 56–69% |
| High | >20% | >20% | ≥70% |

*Note*. MCAS = Massachusetts Comprehensive Assessment System.

aAverage percentage in English language arts and mathematics.

Figure 4. Districts by Proportion of Special Populations (*N* = 35)

Figure 5. Districts by MCAS Achievement (*N* = 35)

## Analysis

During initial contact, the researchers identified an appropriate curriculum, assessment, or accountability administrator at the district and school level for participation in a 30-minute telephone interview. Of the 40 districts invited to participate in Phase 1, 35 did so for a response rate of 88 percent.

Detailed notes were taken during interviewing, and if allowed by the respondent, the interviews were also recorded at the discretion of the interviewer. Notes were compiled and subsequently used to complete a summary report for each participating district that incorporates the district- and school-level data collected. AIR researchers then looked for patterns and themes across districts to determine the most important findings. Districts were also disaggregated based on MCAS 2014 proficiency level, urbanicity, region, and other district characteristics of interest to explore additional themes, differences, and similarities across subsets of districts.

## Data Limitations

Data collection in Phase 1 was designed to collect a brief account of the assessment program in each district, which contributed to the following data limitations:

* Responses to questions about the time required to administer state- and district-mandated assessments varied considerably, reducing our ability to make valid claims about test administration time. The wide range of responses, from statements of the minutes a hypothetical student might need to complete a specific assessment to descriptions of broad testing “seasons,” suggest complexity in thinking about assessment administration that cannot be captured in condensed telephone interviews.
* Variation in the level of detail and specificity provided by respondents suggests that in some districts, multiple personnel might handle different aspects of a district’s assessment system, providing individuals with only part of the picture.

# Findings

## Required District Assessments

Table 2 provides an overview of the required assessments for all participating districts. Across all participating districts, the average number of assessments required by districts is 6.7. Because most assessments are administered multiple times per year, exploring the number of testing sessions students experience per year could be informative. The average number of district-required testing sessions in the sample is 19.

Table 2. District Assessments Across All Responding Districts (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Number of Districts | Average Number of Required District Assessments | Average Number of District-Required Testing Sessions |
| Sample of districts | 35 | 6.7 | 19.1 |

Because the averages in Table 2 aggregate a large amount of data, the average number of district required assessments and testing sessions are shown in Table 3 by grade span. Not all districts in the sample are K–12, so the number of districts serving each grade span is shown. Typically, districts will require two to four assessments per grade span per year and will administer each assessment two or three times per year. The range of assessment approaches includes districts mandating benchmark assessments every five weeks to districts who do not mandate assessments. At elementary and high school levels, districts commonly use assessments targeted at subgroups by grade.

Table 3. District Assessments by Grade Span (*N* = 35)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Number of Districts | Number of Each Grade Span Across Districts | Average Number of Required District Assessments by Grade Span | Number of District-Required Testing Sessions per Grade Span |
| Sample of districts | 35 | ES: 31MS: 29HS: 26 | ES: 2.9MS: 2.6HS: 1.9 | ES: 8.3MS: 7.5HS: 5.6 |

*Note.* ES = elementary school; HS = high school; MS = middle school.

Table 3 shows that the average elementary school student in the sample is required to take 2.9 assessments by their district, in 8.3 separate testing sessions. The table shows also that elementary school students are required by their districts to take slightly more assessments than middle school students. Similarly, middle school students, on average, take more district-required assessments than high school students.

The average number of testing sessions experienced by students per year is illuminating, and an estimate of the average amount of time individual students could spend taking tests would be more so. Unfortunately, estimating the length of testing sessions experienced by students is not possible because the packaged assessments named most frequently by respondents, including Northwest Evaluation Association’s Measures of Academic Progress (MAP) series, Standardized Testing and Reporting assessments, Scholastic Reading Inventory, Fountas and Pinnell (F&P), and Dynamic Indicators of Basic Early Literacy Skills, are not timed. Although several of these companies provide average administration times, the assessments included in the set vary widely in their design, from a few minutes for a student taking an F&P early literacy assessment, to a 60- to 75-minute MAP Goals Survey. These characteristics of commonly used assessments present variables that interfere with accurate estimation.

## Required District Assessments by District Characteristics

As described previously, the districts were sampled by relevant characteristics, accountability level, urbanicity, charter status, region, and MCAS achievement level. In this section, the distribution of the 35 participating districts is explored across each characteristic by the average number of district required assessments and the average number of district-required testing sessions.

**Accountability Level.** As shown in Table 4, Level 4 districts seem to require the lowest average number of assessments, but the small number of these districts in the sample might reduce the validity of this finding. Level 1 districts require the next-lowest average, suggesting that higher numbers of required assessments are clustered in districts at Levels 2 and 3.

Table 4. District Assessments by Accountability Level (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| Accountability Level | Number of Districts | Average Number of District-Required Assessments | Average Number of District-Required Testing Sessions |
| Level 1 | 8 | 5.5 | 14.5 |
| Level 2 | 11 | 8.3 | 22.3 |
| Level 3 | 14 | 6.4 | 19.4 |
| Level 4 | 2 | 5 | 18.5 |

**Urbanicity.** Table 5 shows that the differences between urban and nonurban districts in our sample, in terms of the number of assessments districts require, is not substantial.

Table 5. District Assessments by Urbanicity (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| Urbanicity | Number of Districts | Average Number of District-Required Assessments | Number of District-Required Testing Sessions |
| Urban | 9 | 6.6 | 20.3 |
| Nonurban | 26 | 6.7 | 18.7 |

**Charter Status.** As Table 6 shows, the charter school districts in the study require fewer assessments than traditional districts, on average.

Table 6. District Assessments by Charter Status (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| Charter Status | Number of Districts | Average Number of District-Required Assessments | Number of District-Required Testing Sessions |
| Charter | 7 | 4.3 | 14.7 |
| Noncharter | 28 | 7.3 | 20.2 |

**Region.** Viewing the data by region (Table 7), the lowest average number of district-required assessments is found in the Greater Boston region. The highest average number is in the Northeast region.

Table 7. District Assessments by Region (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| Region | Number of Districts | Average Number of District-Required Assessments | Average Number of District-Required Testing Sessions |
| Berkshires, Southwest | 3 | 5.7 | 17.0 |
| Central | 6 | 7.7 | 23.3 |
| Greater Boston | 9 | 4.4 | 14.3 |
| Northeast | 5 | 10.6 | 28.0 |
| Pioneer Valley | 6 | 5.8 | 16.5 |
| Southeast | 6 | 7.2 | 18.3 |

*Note*. Few districts in The Berkshires and Southwest regions are participating in this study. As adjacent regions in the state, they were combined to protect participants’ anonymity.

**MCAS Achievement.** Considering the data by MCAS achievement level, lower-achieving districts, on average, require fewer assessments. Although medium- and higher-achieving districts seem to require a similar number of assessments, students in medium-achieving districts experience more testing sessions.

Table 8. District Assessments by MCAS 2014 Achievement Level (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| MCAS Achievement Levela | Number of Districts | Average Number of District-Required Assessments | Number of District-Required Testing Sessions |
| Low (40–55%) | 10 | 5.2 | 18 |
| Medium (56–69%) | 14 | 7.3 | 21.6 |
| High (≥70%) | 11 | 7.3 | 17 |

aAverage rating of “proficient” or higher in English language arts and mathematics.

## Preparation for State Tests

Districts and school respondents were asked how much time is devoted to preparing students to take state-mandated assessments. The answers were fairly consistent between districts and schools. When responses differed, preference was given to responses provided by school respondents, based on the assumption that they are more informed about time devoted to practicing for state assessments in their schools. Reports of time spent preparing for state-mandated assessments have been ranked in groups of low, medium, and high amounts of time (Table 9). Districts describing their curriculum as aligned so no preparation time is used, or an amount of time shorter than two days, were rated as *low*. Districts describing preparation times between two and five days were ranked as *medium*. Districts describing preparation times longer than five days, or ongoing preparation on a daily or weekly basis, were ranked as *high*. These divisions were chosen as a substantial number of districts clustered around the two- to five-day range, suggesting a possible statewide norm.

While reviewing the tables below, it is important to keep in mind that relatively few districts in the sample, 17 percent, devote high amounts of time practicing or preparing for state tests (Table 9).

Table 9. Preparation Time for State Tests (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Low (no time to two days) | Medium(two to five days) | High(more than five days and ongoing) |
| All districts | *n* | Percentage | *n* | Percentage | *n* | Percentage |
| 15 | 43% | 14 | 40% | 6 | 17% |

Table 9a shows the preparation time for state tests by accountability level. The data suggest that Level 1 districts tend to devote low amounts of time to test preparation. Districts at Levels 2 and 3 tend to be ranked medium, devoting two to five days to test preparation. Although Level 4 districts in the sample are ranked at medium and high amounts of test preparation time, the small number of districts in this group might not be strongly representative of statewide patterns.

Table 9a. Preparation Time for State Tests by Accountability Level (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| Accountability Level | Low (no time to two days) | Medium(two to five days) | High(more than five days and ongoing) |
| Level 1 | 7 | 1 | 0 |
| Level 2 | 3 | 6 | 2 |
| Level 3 | 5 | 6 | 3 |
| Level 4 | 0 | 1 | 1 |

Table 9b shows preparation time for state tests by region. Smaller numbers in each cell may reduce the reliability of the data, but one notable point is the distribution of Greater Boston districts with most in the low range of state test preparation time, and no districts ranked as high.

Table 9b. Preparation Time for State Tests by Region (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| Region | Low (no time to two days) | Medium(two to five days) | High(more than five days and ongoing) |
| Berkshires, Southwest | 2 | 0 | 1 |
| Central | 1 | 2 | 3 |
| Greater Boston | 6 | 3 | 0 |
| Northeast | 2 | 3 | 0 |
| Pioneer Valley | 2 | 2 | 2 |
| Southeast | 2 | 4 | 0 |

Table 9c shows preparation time for state tests by urbanicity. Notable in this table is the distribution of districts ranked low, medium, and high is similar for both categories of districts with an even division between low and medium rankings, and a smaller number of districts that are ranked high. This pattern reflects the statewide distribution of preparation time for tests, suggesting that urban and nonurban districts are equally likely to spend low, medium, or high amounts of time preparing their students for state tests.

Table 9c. Preparation Time for State Tests by Urbanicity (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| Urbanicity | Low (no time to two days) | Medium(two to five days) | High(more than five days and ongoing) |
| Urban | 3 | 4 | 2 |
| Nonurban | 12 | 10 | 4 |

Table 9d shows preparation time for state tests by district type. While acknowledging a substantial difference in the number of districts in each category, nearly all charter districts report devoting low amounts of time to test preparation, whereas traditional districts are distributed across the three groups. This may suggest that charter districts generally spend less time on test preparation than traditional districts.

Table 9d. Preparation Time for State Tests by District Type (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| School Type | Low (no time to two days) | Medium(two to five days) | High(more than five days and ongoing) |
| Charter | 6 | 1 | 0 |
| Traditional | 9 | 13 | 6 |

Table 9e shows preparation time for state tests by MCAS achievement level. Using AIR’s categorization of MCAS achievement into groups of low, medium, and high, the table suggests that the districts ranked high in MCAS achievement tend to spend low to medium amounts of time preparing for state tests. Districts ranked low and medium in MCAS achievement show rankings for test preparation time that are distributed across the three preparation groups, with medium-proficient districts distributed more heavily on the low end of preparation time, and districts ranked low in MCAS achievement distributed more heavily across medium and high categories for test preparation time. Although these patterns are visible, they are not particularly robust given the small differences in categories and the small numbers of districts in each category.

Table 9e. Preparation Time for State Tests by MCAS Achievement (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| MCAS Achievement | Low (no time to two days) | Medium(two to five days) | High(more than five days and ongoing) |
| Low (40–55%) | 2 | 5 | 3 |
| Medium (56–69%) | 7 | 4 | 3 |
| High (≥70%) | 6 | 5 | 0 |

Table 9f shows the preparation time for state tests by percentage of ELLs in participating districts. Using AIR’s categorization of the percentage of ELL students into groups of low, medium, and high, the data are distributed across the preparation time rankings in a manner that is consistent with statewide patterns established in Table 9, suggesting that districts are as likely to devote low, medium, or high amounts of time to test preparation, regardless of the percentage of ELL students in the district.

Table 9f. Preparation Time for State Tests by Percentage of ELLs (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| Percentage of ELLs | Low (no time to two days) | Medium(two to five days) | High(more than five days and ongoing) |
| Low (0–10%) | 5 | 7 | 4 |
| Medium (11–20%) | 3 | 5 | 1 |
| High (>20%) | 3 | 5 | 1 |

Table 9g shows the preparation time for state tests by the percentage of students with disabilities (SWDs) in participating districts. Using AIR’s categorization of the percentage of SWDs into groups of low, medium, and high, the middle group reflects the statewide pattern of an even distribution between low and medium rankings of time devoted to test preparation, and a smaller number of districts ranked as high in test preparation time. The districts ranked low and high in terms of percentage of SWDs are each distributed toward the low end of the preparation time rankings, but they are small in number and not necessarily representative of statewide trends.

Table 9g. Preparation Time for State Tests by Percentage of SWDs (*N* = 35)

|  |  |  |  |
| --- | --- | --- | --- |
| Percentage of SWDs | Low (no time to two days) | Medium(two to five days) | High(more than five days and ongoing) |
| Low (0–10%) | 2 | 0 | 0 |
| Medium (11–20%) | 11 | 11 | 6 |
| High (>20%) | 2 | 3 | 0 |

## Test Scheduling

Participating districts uniformly talked about the general lack of flexibility in scheduling state-required tests, which must be administered within a period of time mandated by the state. Often, the decision about the exact dates of testing within the specified window is made by district administrators in consultation with building administrators at each grade span, instructional coaches, and school leadership teams. Other considerations for scheduling state-mandated tests include the following:

* Availability of district and school staff
* Availability of computers
* Accommodations for ELLs and SWDs
* Academic readiness of students to take the tests

Although district-required tests are also associated with a window for test taking, administration of these tests is perceived as being more flexible by district and school respondents, particularly so for tests that monitor student progress on learning goals or for formative purpose such as informing instruction (less so with benchmarking assessments). This is largely because teachers are often involved in deciding when these district-level assessments are administered in their classrooms, and their decisions are driven by where they are in teaching standardized curricula or units of study. Other considerations for scheduling district-required tests include the following:

* Proximity to other tests, particularly state-mandated tests, and the need to balance instruction and assessment
* Academic calendar and where the quarter or semester breaks are
* Curriculum maps developed by the school or district
* Scheduled district or school events
* Timing of tests to effectively monitor student progress/growth
* Availability of school staff

Among participating Phase 1 districts, those that described a more thoughtful approach to test scheduling (i.e., took into consideration more of the factors mentioned in the preceding list), the majority were either in the medium or high category of MCAS achievement.

## District Support

Based on Phase 1 interview data, district support for test preparation and administration as well as analyses of assessment data are generally provided in several common forms that include the following:

* Training events at the district, regional, or state level
* Professional learning opportunities within schools or at the district level
* Testing guidance, manuals, materials, and technology (computers as well as specialized data services such as TestWiz)
* Assessment or data expertise (specialist or team) at the district or school level

With the data collected, it is not possible to differentiate districts based on the level of testing support they provide to schools. All schools described some combination of the supports listed above, and it was not feasible to discern the effectiveness of the supports provided.

## Maximizing Instructional Time

In addition to looking at the process of scheduling district and state assessments, the study is also interested in examining the extent to which and how districts maintain or protect instructional time around state-mandated tests. Several factors that could impact instructional time were considered in Phase 1:

1. The amount of time districts and schools allot to state testing (within the window set by the state)
2. The preparation time for state tests
3. The use of student time on testing days before and after testing sessions

Analyses of Phase 1 interview data on these three conditions were inconclusive but provided preliminary evidence that some districts may be more attentive to these factors and to protecting instructional time overall. These districts will be studied in greater depth in Phase 2.

# Summary

Based on brief district- and school-level interviews conducted in a sample of 35 representative districts from across the state, the findings regarding required district assessments and preparation time for state tests are presented next.

## Required District Assessments

Across all participating districts, the average number of assessment required by districts is 6.7, and the average number of district-required testing sessions in the sample is 19.1. Districts typically require two to four assessments per grade span per year, and each assessment will be administered two or three times per year. Based on interviews, elementary school students are required by their districts to take slightly more assessments than middle school students, and middle school students, on average, take more district-required assessments than high school students.

When districts were disaggregated by accountability level, Level 4 districts seem to require the lowest average number of assessments, whereas Level 2 required the highest. This coincides with findings by MCAS achievement levels. That is, lower-achieving districts, on average, require fewer assessments, although students in districts at the middle level of MCAS achievement seem to experience more testing sessions.

When the geographic region is taken into consideration, the lowest average number of district-required assessments is found in the Greater Boston region, and the highest average number is in the Northeast region. Districts in this region require nearly three times as many assessments as the counterparts in the Greater Boston region.

Participating districts also reported that students spend virtually no time preparing for or practicing district-required assessments as such testing is generally integrated into instruction and curricula.

## Preparation for State Assessments

Districts and school respondents were asked how much time is devoted to preparing students to take state-mandated assessments. Most districts fell in the low or medium range (up to five days of preparation time). Highlights from these data include:

* Most Level 1 districts spend no more than two days preparing for state tests, while districts in Level 2 and Level 3 were more spread out, with many allocating 2–5 days on average preparing their students for state tests.
* Most districts in the Greater Boston region do not spend more than two days preparing their students for state tests, whereas districts in other regions were more varied. Districts in the Central region were more likely to be in the medium-high range of test preparation, and districts in the Northwest and Southeast regions were in the low-medium range.
* Urban and nonurban districts are equally likely to spend low, medium, or high amounts of time on test preparation
* Traditional districts, on average, spend more time preparing their students for statewide testing than charter districts.
* Districts ranked high in MCAS achievement tend to spend lower amounts of time preparing for state tests. Districts in the low and medium categories of MCAS achievement were more spread out, with medium-achievement districts distributed more heavily on the low end of preparation time, and districts ranked low in MCAS achievement more likely to be in the medium to high range.

## Test Scheduling

Participating districts uniformly talked about the general lack of flexibility in scheduling state-required tests, which must be administered within a period of time mandated by the state. Often, the decision about the exact dates of testing within the specified window is made by district administrators in consultation with school leaders at each grade span.

Although a window for test taking is also associated with many district-required tests, the administration of these tests is perceived as being more flexible by district and school respondents. This is largely because teachers are often involved in deciding when these district-level assessments are administered in their classrooms, and their decisions are driven by where they are in core concepts or units of study.

## District Support

District support for test preparation, test administration, and analysis of assessment data is generally provided through opportunities to attend structured training events at the district, regional or state level; funding for professional learning within schools or at the district level; testing guidance, manuals, materials, and technology; and assessment or data expertise at the district or school level.

## Phase 2

The information collected in Phase 1 will inform the selection of a smaller sample of districts for Phase 2, which entails a closer examination of each district’s assessment program and practices. Phase 2 also provides a chance for participating districts to reflect on the strengths and weaknesses of their required assessments and to share promising practices or strategies related to the preparation for and timing of state and local assessments so that instructional time is not compromised.