Measuring Implementation in Massachusetts Expanded Learning Time Schools

May 2013

Prepared for:
Massachusetts Department of Elementary and Secondary Education (ESE)
75 Pleasant Street, 5th Floor
Malden, MA 02148

Submitted by:
Abt Associates Inc.
Beth Gamse
Amy Checkoway
Catherine Darrow
55 Wheeler Street
Cambridge, MA 02138
Identifying programs that demonstrably improve educational experiences and outcomes for students is essential for reducing educational disparities. Yet over the past several decades, few interventions at any level of the educational system have produced the desired positive effects at scale. When small scale interventions implemented by program developers appear promising, yet falter when implemented more widely, policymakers and researchers alike seek to understand why the results differ across settings. Increased understanding of variation in outcomes has led to more systematic examination of program adoption and implementation elements (Fixsen, Blasé, Naoom, & Wallace, 2009).

This paper describes methodical efforts to measure the implementation of an initiative designed to increase instructional time for students in a sample of Massachusetts schools. Although efforts to implement expanded instructional time had occurred in individual schools within Massachusetts (either as stand-alone initiatives or as part of larger school restructuring reforms), the 2005 establishment of a state-level initiative represented the first example of larger-scale implementation.

Increasing the amount of instructional time is often perceived as a promising initiative for improving student achievement, and there is some research evidence of a positive relationship between increased instructional time and academic outcomes (Angrist, Dynarski, Kane, Pathak, & Walters, 2010; Atwater & Morre, 1991; Doran & Drury, 2002; Educational Policy Institute, 2005; Gallagher & Ross, 2005; Hoxby, Murarka, & Kang, 2009; Tuttle, The, Nichols-Barrer, Gill, & Gleason, 2010). However, there is, as yet, limited rigorous research that has demonstrated positive impacts of expanded instructional time on student outcomes (Patall et al., 2010), and further, existing studies have focused primarily on models that examine the combined effects of an extended school day with other reforms (Redd et al., 2012).

The implementation-focused study that is discussed in this paper was part of a larger evaluation of the Massachusetts Expanded Learning Time (ELT) initiative conducted by Abt Associates that provides further evidence about ELT implementation, outcomes, and the relationship between them. The evaluation was supported by a grant from the Institute of Education Sciences in the U.S. Department of Education to the Massachusetts Department of Elementary and Secondary Education (ESE); the initiative (and its evaluation) also received some support from additional sources including ESE and initial funds from Massachusetts 2020.¹

The paper begins by briefly summarizing the value and purposes of measuring implementation. We then present a detailed description of an index developed to examine implementation of the Massachusetts ELT initiative, followed by a discussion of the implications of this work for future efforts.

**Why Measure Implementation?**

Especially when a program is in its earlier stages, learning about differences in how core elements are delivered is helpful for clarifying program guidance and refining activities, as well as for understanding whether and how local circumstances shape program delivery. Many program developers have a theory of action, or a model of change, about the program elements that they believe will lead to specific outcomes for participating individuals and/or organizations such as schools. Monitoring and documenting a

¹ The IES grant award number for this study is R305A080372.
program’s resources, training, support, and practices can be used for planning expansion or replication efforts as well as provide information about if the intervention was implemented as intended. Program monitoring can reveal cases of high implementation as well as occurrences of program drift (Bond, Evans, Salyers, & Kim, 2000; Mowbray, Holter, Teague, & Bybee, 2003).

When evidence emerges that an initiative was not implemented as intended, technical assistance providers, coaches, and other support staff can target those facing implementation challenges to refocus efforts. Additional assistance can serve multiple purposes, including strengthening aspects of problematic implementation, ensuring greater consistency in multi-site implementation efforts, and helping program developers, funders, and other stakeholders articulate which supports are necessary for meaningful implementation of program components. In the case of the Massachusetts ELT initiative, earlier data about implementation variation contributed to more specific and detailed guidance from the Massachusetts Department of Elementary and Secondary Education as well as changes in the nature of technical assistance provided to ELT schools.

Further, measuring implementation fidelity represents one strategy for examining if outcomes of interest are indeed related to the delivery of an intervention, and therefore can help establish internal validity. In other words, researchers can only confirm that outcomes can be attributed to a program if the critical components of that program were indeed delivered as intended; evidence of high levels of fidelity is a necessary (though not sufficient) ingredient in establishing the relationship between elements of the intervention and the resulting outcome(s) of interest (Gresham, Gansle, & Noell, 1993).

Fidelity measures can provide information about whether and how critical programmatic elements are delivered in the targeted settings (in this case, schools) as well as those settings (and their participants) serving as comparisons. Researchers can assess whether the implementation elements were decidedly different for one or the other settings/groups. For the ELT initiative, implementation measures examined if ELT schools were implementing the requisite elements and then if and how the time in the school day was systematically differently than for students who attended other (non-ELT) schools. If differences are not verified, any subsequent differences in outcomes cannot justifiably be attributed to the intervention.

In sum, given the importance of capturing the level of implementation fidelity accurately, research needs to focus on distinctive program components theorized to produce desired outcomes. Such measures can capitalize upon data from multiple sources and across multiple time points. However, researchers should also balance the availability of rich data about various aspects of implementation with the potentially daunting task of synthesizing those data to answer questions about how the program was implemented, by whom, and under what circumstances (Century, Rudnick, & Freeman, 2010). Recent research highlights the importance of collecting and analyzing implementation data in program evaluation, whether within academic interventions (Century et al., 2010), drug abuse preventions (Dusenbury, Brannigan, Falco, & Hansen, 2003), and across social and educational programs (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). Yet even as researchers increasingly acknowledge the importance of implementation data, comprehensive guidance about to use implementation data well—including transforming data into a meaningful and systematic index—is still in the process of development and refinement.

### Case of Massachusetts Expanded Learning Time (ELT)

We now turn to describing the development of an implementation index in the context of the Massachusetts Expanded Learning Time (ELT) initiative. The Massachusetts ELT initiative provides state-funded grants to selected schools across multiple districts to increase instructional time by at least
300 hours per academic year. Participating schools receive an additional $1,300 per student to lengthen the day and/or year. The evaluation included the first cohort of 10 ELT schools (in five districts) that received implementation grants for their expanded programs in the fall of 2006, a second cohort of nine schools (and three new districts) that began implementation in the fall of 2007, and nine ELT schools (and four new districts) that began implementation in 2008-09.  

While the specific expectations for how schools should use their additional time have evolved over time, the major objectives remain the same: 1) provide more instructional opportunities in core academic subjects; 2) integrate enrichment opportunities into student learning; and 3) provide educators with increased opportunities to plan instruction and participate in professional development. From the beginning of the Massachusetts ELT initiative, formal technical assistance was provided to most ELT schools from Massachusetts 2020 (Mass 2020), now a state affiliate of the National Center on Time and Learning, and Focus on Results through leadership training sessions and/or ongoing individualized coaching.

After three years of examining implementation and early outcomes of the ELT initiative as successive cohorts of schools began to implement ELT, we recognized that our efforts to describe and understand implementation of a multi-faceted initiative were too limited. We believed that we needed to understand implementation of specific program components more thoroughly, and ultimately explore whether certain implementation factors might be influencing the study’s outcomes findings. This interest was driven in part by the evaluation’s lack of consistent impacts on student achievement outcomes over the course of the study period and the desire to better understand the role of implementation.

As of the 2009-10 school year, the study was collecting interview and survey data about various aspects of increased instructional time (e.g., availability of staff support, professional development and technical assistance in general, satisfaction with the expanded day structure). However, such information was not detailed enough to understand variation in implementation systematically across the ELT schools. In particular, more specific information about student and staff schedules, and about the amount of time available for different types of instruction and/or students was desired. It should be noted that information about time allocations is not straightforward, especially in middle grades, when no two students (or teachers) are likely to have the same schedule or spend the same amount of time on specific subject areas. In addition, schools do not necessarily use the same terminology to identify certain instructional activities (e.g., “electives” versus “specials,” “remediation” versus “academic support”).

As a result of these discussions, Abt Associates developed an implementation index, working collaboratively with ESE and Mass 2020. Our intent was to develop a measure that would incorporate data from multiple sources, especially data about the amount of time allocated to different elements of the ELT initiative. The index was designed to incorporate data from district, principal, and school partner interviews, teacher and student surveys, and master school schedules. The index allowed the study team to describe implementation activities during the 2009-10 year more systematically; it was modified in the subsequent year to reflect changes in expectations of ELT schools as well as changes in data collection instruments.

It is important to acknowledge that the data used to construct the implementation index were collected at various time points during the academic year, and that the index scores for that year were ultimately

---

2 Over the course of the ELT initiative, a handful of schools have restructured (e.g., merged with another school), closed, or exited. Two new ELT schools were funded in 2012-13.
generated in the following fall. Based on review by the study team, ESE, and Mass 2020, the index was revised in 2010 to use in examining data collected in the 2010-11 academic year.

The implementation index was designed to serve four major purposes. One, the evaluation team wanted to learn more systematically whether there were specific aspects of implementation that ELT schools were successfully or having difficulty putting into place; by applying the same implementation measure to all study schools, the evaluation team could describe implementation efforts more consistently. Two, the index was intended to generate information for ESE about specific patterns in implementation, and that information then could help shape program guidance, monitoring and supports. Three, the index was designed to produce systematic data about implementation across participating schools for technical assistance providers to use in customizing subsequent supports according to individual school needs. Fourth, the index was intended to inform exploratory outcome analyses linking outcomes to levels of implementation.

The index was purposefully designed to apply to ELT (i.e., program or treatment) and to non-ELT (i.e., comparison) schools, an important feature given that (1) some districts had begun to expand time in additional schools beyond those funded by the state’s ELT initiative, and (2) some of the core expectations of ELT apply across multiple school reform initiatives, and consequently are not limited to the ELT initiative. As such, the index was used to measure levels of implementation in both ELT and matched comparison schools during 2010-11.

The index builds on five broad foundational principles of the ELT model that had been articulated by Mass 2020. These principles address such elements of ELT as additional time for core instruction, school leadership, time allocation for instruction, and teacher planning/collaboration, and are described in more detail below. The index uses data from interviews with principals as well as from surveys of teachers and students; as such, it provides a snapshot of ELT implementation at a particular point in time. Note that information about school schedules (and time allocation) was collected only for fifth and eighth grade students, and only students in those grades were administered surveys, although all instructional staff in each study school were asked to complete surveys.

The implementation index (see Exhibit 1) includes five criteria based on core expectations for ELT implementation, including:

1) school-wide academic focus;
2) core academics;
3) enrichment activities;
4) teacher leadership and collaboration; and
5) school leadership.

Each criterion includes four levels, ranging from zero (indicating no or little discernable activity), to Level 1 (representing limited evidence that schools have implemented activities related to that criterion), to Level 2 (representing modest evidence that schools have implemented activities related to that criterion), to Level 3 (representing consistent evidence that schools have implemented activities related to that criterion). Each consecutive level assumes that all the conditions outlined in the prior level have been met. Three of the criteria described above (core academics, enrichment, and teacher leadership and
collaboration) are measured by two components: amount of time engaged in the activity and how the time is used.

The index also includes a sixth criterion applicable only to ELT schools, about the level of support for ELT within schools and districts. This criterion was not included in calculations of the overall score. (See appendix for more detail on the criteria.)

**Applying the Implementation Index**

The specific thresholds for components of the index were determined using an iterative process, and decisions were tempered by expectations of what might be reasonable for schools to achieve after two, three, four, or five years of implementation (corresponding both to the school year for which data were examined and to the staggered number of years each cohort had been implementing ELT). Based on data from surveys, interviews, and school schedules, schools were assigned a criterion-specific score that ranged from zero (insufficient evidence of achieving Level 1) to three (corresponding to Level 3 of implementation). The process used to define thresholds that would discern differences between levels of implementation for each of the core components required both (1) substantive discussions about what behaviors and perceptions would characterize minimal, moderate, or consistent implementation of each ELT expectation, and (2) empirical discussions based on actual distributions of survey and interview responses. Two criteria each had two components (amount of time allocated and how time was used), and therefore generated two scores each. The index included eight separate scores, and a maximum total of 24 points. A zero score did not necessarily mean that a school did not implement a particular criterion at all, but rather that it did not meet the threshold necessary for Level 1. Each criterion was weighted equally.

ELT schools, on average, scored higher than comparison schools on six of eight criteria. Comparison schools scored, on average, higher than ELT schools on the school-wide academic focus and school leadership criteria. Comparison school scores were generally between Level 0 and Level 1 on how core academic time was used, enrichment time, and teacher leadership and collaboration time criteria (averages of 0.47, 0.44, and 0.53, respectively). ELT school scores were generally between Level 0 and Level 1 on the school-wide academic focus criterion (0.78). Differences were not tested for statistical significance, so the meaning of any differences between schools or groups of schools is unknown.

**Discussion**

We noted above that the index was designed to address multiple purposes, including learning whether: (1) certain aspects of implementation were more/less difficult for schools; (2) patterns in implementation could inform program guidance and support from ESE; (3) data about implementation across participating schools might help the technical assistance providers’ efforts to serve schools; and (4) whether variation in implementation was associated with variation in outcomes.

The index yielded valuable information that served all four purposes to some degree. We learned that the structural elements of expanding the school day (i.e., adding the required number of minutes each day) appeared to be easier for schools to implement than were changes in how that additional time was used, for example. ESE refined the content of performance agreements with all participating schools, based on its own review processes, along with the overall implementation index scores, and began to customize the agreements to reflect school-specific achievements and goals. The technical assistance providers adapted their efforts (in the next school year) to strengthen implementation across schools, and we used the index
scores in exploratory analyses, although preliminary analyses showed that variation in implementation was not predictive of variation in student achievement.

Beyond its utility in serving various instrumental purposes, the index also provided systematic corroboration for the study’s persistent findings about the overwhelmingly important role of implementation. In fact, variation in schools’ levels of implementation was one of the major takeaway findings from the study—across both program and comparison schools, across middle and elementary schools, and across all study schools. The consistent variability led the study team to consider other possible explanations and influences, ranging from the initial selection process, to the staggered cohort approach, to the autonomy of school-by-school decision-making.

Of course, the index also had limitations, and was neither designed nor capable of addressing all possible questions or needs. First, the index was primarily constructed to reflect the presence or absence of structural elements. The development of the index’s content during its first year relied, out of necessity, upon the data that had already been collected. Although the study team revised data collection protocols in part due to a desire to fine-tune the index, those revisions applied only to the study’s final year of data collection. And while the revised instruments allowed us to describe elements of implementation in greater detail, the focus of the index continued to be on the level versus the quality of program component implementation. Future efforts would certainly benefit from focus on both structural elements (i.e., presence/absence of key implementation components) as well as indicators of program quality, which could then inform program accountability and monitoring activities, technical assistance efforts, and evaluation analyses focused on implementation variability.

Second, the development and use of the index occurred within the context of the study’s pre-existing agreements with study schools to treat all data as confidential. That decision meant keeping schools’ identities confidential, even to ESE. This limited the state’s capacity to engage in programmatic follow-up with specific schools about particular implementation challenges. Future evaluation efforts might weigh the relative benefits and challenges of naming individual schools for certain evaluation purposes, even when data are aggregated to the school levels, to be able to identify schools with particular implementation successes or challenges. Consideration should be given to any effects that school-level reporting could have on the validity of data. Confidentiality agreements that include clear assurances about the purpose of school-level aggregation and identification of schools (e.g. for program improvement) may help assuage concerns about how data might be used.

In sum, despite some limitations, the development of the implementation index enabled us to integrate information from multiple data sources into one measure, to describe variation in implementation efforts, and to recognize how important understanding that variation was. It also allowed us to apply a measure that could be used to explore relationships between levels of implementation and student outcomes. Both of these results have proven useful, and the use of similar indices and tools should prove beneficial to other researchers, program developers, and policymakers who seek to better understand implementation as well as understand the relationship between implementation and desired outcomes.
References


Appendix: Massachusetts Implementation Index Criteria, 2010-11

Criterion 1: School-Wide Academic Focus
The academic focus criterion addresses whether a school-wide academic focus is present and consistently identified and disseminated (Level 1), whether the focus is linked to instruction (Level 2), and whether the focus is regularly monitored and used to adjust instructional practices (Level 3).

Criterion 2: Core Academics in Target Grades
The time component of the core academics criterion addresses whether students in target grades receive at least 300 minutes of ELA and at least 250 minutes of math instruction per week (Level 1), at least 350 minutes of ELA instruction, at least 300 minutes of math instruction, and at least 150 minutes of science or social studies instruction per week (Level 2), and at least 200 minutes of both science and social studies instruction in addition to Level 2 ELA and math thresholds (Level 3). The “how” component addresses teacher support for improvements to academic instruction and student receipt of at least 50 minutes of dedicated academic support time per week (Level 1), student receipt of at least 100 minutes of academic support time per week and whether specific particular instructional practices are used (Level 2), and student receipt of at least 150 minutes of academic support time in addition to a higher threshold for the use of particular instructional strategies (Level 3).

Criterion 3: Enrichment Activities in Target Grades
The time component of the enrichment criterion addresses whether the school offers at least 45 minutes of enrichment activities and 45 minutes of specials per week (Level 1), at least 90 minutes of each per week (Level 2), and at least 135 minutes of each (Level 3). The “how” component addresses student access to enrichment (Level 1), student choice about and school integration of enrichment (Level 2), and the extent to which enrichment activities are influenced by the school’s academic focus, connected to grade level standards and/or curriculum frameworks, and perceived by teachers to be of high quality (Level 3).

Criterion 4: Teacher Leadership and Collaboration
The time component of the teacher leadership and collaboration criterion addresses whether the majority of teachers attend collaborative planning meetings at least twice monthly for at least 45 minutes (Level 1), at least once per week and whether/how frequently collaborative planning time has been supplanted by other unrelated activities (Level 2), and more than once per week (Level 3). The “how” component addresses the structure and content of collaborative planning time (Level 1), the culture of collaboration (Level 2), and opportunities for teacher leadership in a school (Level 3).

Criterion 5: School Leadership
The school leadership criterion addresses teacher perceptions of school leadership direction setting (Level 1), school leadership involvement in instructional practice and staff development (Level 2), and school leadership focus on strengthening school culture (Level 3).

Criterion 6: ELT Support
A subset of indicators of ELT implementation applicable only to funded ELT schools comprises the ELT support criterion. This criterion addresses the extent to which teachers are supportive of ELT and report that school leadership sets high expectations for ELT (Level 1), whether and how the leadership team involves others in designing and implementing ELT changes and teacher perceptions of district support for ELT implementation (Level 2), and if the majority of teachers report satisfaction with staffing provided to cover the entire school day (Level 3).