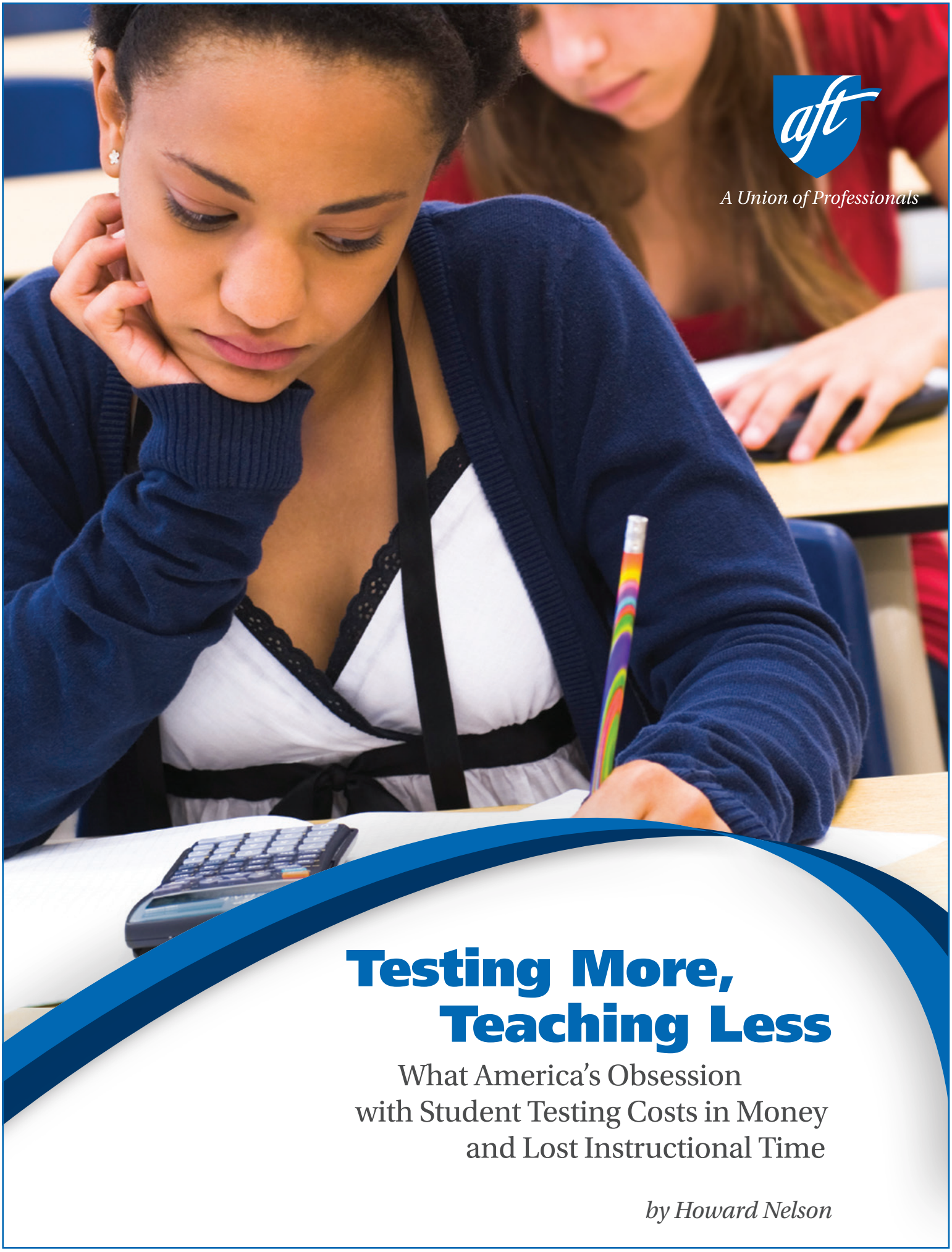


A Union of Professionals

Testing More, Teaching Less

What America's Obsession with Student Testing Costs in Money and Lost Instructional Time

by Howard Nelson

A Union of Professionals

Testing More, Teaching Less

What America's Obsession with Student Testing Costs in Money and Lost Instructional Time

by Howard Nelson

A Union of Professionals

OUR MISSION
The **American Federation of Teachers** is a union of professionals that champions fairness; democracy; economic opportunity; and high-quality public education, healthcare and public services for our students, their families and our communities. We are committed to advancing these principles through community engagement, organizing, collective bargaining and political activism, and especially through the work our members do.

Introduction

All children deserve a rich, meaningful public education that prepares them for the opportunities, responsibilities and challenges that await them as contributing members of a democratic society and a global economy. That vision should be as true for students in Birmingham or the South Bronx as it is for those in Beverly Hills. And it's why the AFT and our affiliates have been advocates of clear, common, core curricular standards for more than 20 years, and why we strongly support the Common Core State Standards (CCSS) for mathematics and English language arts and other career- and college-ready standards efforts today. But the deeper learning we aspire to has too often been a rhetorical aspiration—not accompanied by the supports needed to make implementation a reality—and eclipsed by the misuse and overuse of standardized assessments required by policymakers fixated on accountability above all else.

The coupling of state standards and assessments to measure and report student and school performance under the No Child Left Behind Act narrowed curricula across the country. Despite a laudatory goal of shining the light on student needs, it took us in another direction, away from valuing the essential skills of persistence, critical thinking and collaboration. Instead of resulting in the standards-based public education system our nation and our children deserve, the current standardized test-based accountability system has left classroom teachers almost solely responsible for the performance of students and schools. Many districts piled on, adding a plethora of other standardized tests to benchmark student performance levels, measure progress towards a state's standardized test, or layer on requirements for promotion or graduation.

Educators know the necessity of gauging student learning—they use various assessment techniques throughout the school day. And we support the proper use of standardized testing and sensible accountability measures. Educators, parents and others have joined the AFT's efforts to restore the balance between teaching and testing, most recently through our "Learning Is More Than a Test Score" campaign. The current test-and-punish accountability system has squeezed out vital parts of the curriculum that are not subjected to accountability testing, sacrificed student learning time to testing and test preparation, and forced teachers—particularly those teaching our most vulnerable students—to focus their attention on students achieving just below the passing score. That is not what countries with high-performing education systems do, and it is not what the United States should do.

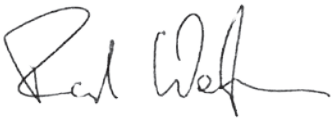
Last summer, delegates to the AFT convention went on record in support of testing that informs, rather than impedes, teaching and learning, and in favor of studies that shed light on the real costs of testing. *Testing More, Teaching Less* is part of delivering on our commitment to provide guidelines, studies and other helpful information to our members and the broader public about the nature, amount and costs of student assessments. Many other stakeholders have voiced their concerns about the impact of standardized tests and have taken action to curtail overtesting and its consequences. In Texas, lawmakers cut the number of high school end-of-course exams required for graduation from 15 to five, and eliminated the requirement that results would count for 15 percent of a student's overall grade. The Orchard Park Central School District Board of Education in New York took a stand with a resolution proposing that this year's state assessments be used for "measuring the state's progress in introducing the Common Core Learning Standards rather than for measuring student performance or educator effectiveness." Lawmakers in New Mexico called for an analysis of the cost, both in instructional time and money, of all student assessments. And just this month, the *New York Times* ended a strongly worded editorial about the dangers of "testing mania" with a call for the country to "reconsider its obsession with testing, which can make education worse, not better."

We're at a point where the adoption and implementation of the Common Core State Standards for the majority of the nation's students should be deepening the potential for all students to learn what they need to be college- or career-ready upon gradu-

ation. Instead, public policy and politics have put the testing cart before the standards-and-curriculum horse. As educators, we know that any meaningful standards need more than assessments to succeed. Resources for improved instruction are necessary, in addition to aligned curriculum, professional development, and time and tools for teachers to learn and collaborate. Unfortunately, the majority of teachers recently polled by the AFT say that this is not happening—or it is happening without the voice and views of teachers. Yet, states are moving forward with assessments that have consequences for students, teachers and schools. That's why I called for a moratorium on high-stakes consequences associated with new standards-based assessments. The U.S. Department of Education heard and responded to the voices of teachers who haven't had enough time or support to translate these standards into classroom practice, allowing states to ask for additional time before using outcomes of new assessments based on the CCSS to evaluate teachers.

In this climate, it is more important than ever that we look at testing and its impacts in a more informed way. This analysis aims to illuminate with data what many educators know from experience, parents learn from their children, and students feel firsthand—testing has spiraled out of control, and the related costs are unacceptably high and are taking their educational toll on students, teachers, principals and schools. Our study examines that very concern: the total cost of testing, including the direct (financial) costs of tests and the cost of instructional time lost to testing and test preparation associated with two districts' testing schedules. This is an illustration of the kind of information we as professionals and as a public need in order to ask and answer the right questions about the role of testing in our schools. How much do we spend? What do we get in return? What do we give up as a consequence? What are the best ways to use precious instructional time and resources so all our students have the content, critical thinking, problem-solving and persistence they need to succeed in school and in life? *Testing More, Teaching Less* doesn't provide all the answers. But it makes clear that the current testing environment is inhospitable to the knowledge, skills, and abilities we aspire to develop in students—and it offers some concrete recommendations for correcting the dangerous course we are on.

I would like to gratefully acknowledge thoughtful reviews of an earlier version of this analysis from Jesse Rothstein of the University of California-Berkeley, Elena Silva of the Carnegie Foundation for the Advancement of Teaching, and a reviewer who prefers to remain anonymous.



Randi Weingarten
President, American Federation of Teachers
July 2013

Abstract

Based on a detailed grade-by-grade analysis of the testing calendars for two mid-size urban school districts, and the applied research from other studies of state mandated testing, our study found that the time students spend taking tests ranged from 20 to 50 hours per year in heavily tested grades. In addition, students can spend 60 to more than 110 hours per year in test prep in high-stakes testing grades. Including the cost of lost instructional time (at $6.15 per hour, equivalent to the per-student cost of adding one hour to the school day), the estimated annual testing cost per pupil ranged from $700 to more than $1,000 per pupil in several grades that had the most testing. If testing were abandoned altogether, one school district in this study could add from 20 to 40 minutes of instruction to each school day for most grades. The other school district would be able to add almost an entire class period to the school day for grades 6-11. Additionally, in most grades, more than $100 per test-taker could be reallocated to purchase instructional programs, technology or to buy better tests. Cutting testing time and costs in half still would yield significant gains to the instructional day, and free up enough dollars in the budget that could fund tests that are better aligned to the standards and produce useful information for teachers, students and parents.

Executive Summary

OBJECTIVES

Students are engaged in various types of testing each year. The purpose and quality of such testing, the time spent taking and administering tests and the usefulness of the results are ongoing topics of discussion among students, parents, educators and policymakers. Advocates for more testing implicitly assume that more testing improves student achievement and that the benefits of more testing outweigh the additional costs, which they typically view only as the money paid to testing companies. Opponents of excessive testing claim that schools have sacrificed learning time in favor of testing and test preparation, reduced learning time for non-test subjects, and focused time and attention on the "bubble kids" (students whose scores are clustered right below the proficiency cut score) at the expense of developing every student's full potential.

To get a complete picture of the resources now devoted to testing in the United States, our study documents the types of assessments, the number of assessments and the number of times each test is administered annually, as well as the associated test-taking time and the direct budgetary cost of such tests. Our analysis encourages policymakers to judge the benefits of current testing policy relative to budgetary costs and alternative uses of student and teacher time, such as more instruction and more attention to non-test subjects.

Although more comprehensive than most other studies of the cost of testing, our study excludes many testing costs that could have a significant impact, such as a teacher's non-classroom time preparing for testing, the cost of test prep materials, the extra hours spent with special needs and ELL (English language learner) students due to testing accommodations, the cost of tests specifically administered only to special education and ELL students, lost services from reading and special education teachers when they administer or proctor tests, cost of data/instructional coaches and teacher time lost to data reporting and analysis activities, hardware and technology costs attributable to testing, time spent on assessing and grading homework and classroom tests, and the costs of tutoring and summer school linked to high-stakes testing. Outside the scope of our study are the hours students spend taking quizzes or teacher-made tests, and the time teachers spend grading tests and homework. In Chicago, for example, teachers report spending 32 minutes per day, "assessing students' work during contractual hours" and 22 minutes a day "giving curriculum subject assessments (tests, quizzes, etc.)" (Bruno et al., 2012).

METHODOLOGY

To gather this information, the AFT collected the assessment inventory and testing calendar from two, medium-size urban school districts with the pseudonyms Midwestern School District and Eastern School District. In both districts, the AFT had very good access to the assessment inventories as well as time and cost data. Unlike a case study, a two-district analysis recognizes variety in assessment practices. One district has more testing than the other and the states in which they are located also reflect a big difference in state-mandated testing time and test quality. Although neither state was among the highest or lowest spending, one state spent twice as much on state testing as the other (Chingos, 2012).

The direct budget costs of the tests and logistical support for testing were estimated, as well as the time needed for students to take and prepare for the tests. Time and cost data were provided in district documentation or came from Internet searches of commercial test publishers. The information is presented by grade because the cost and lost instructional time vary greatly by grade; students and teachers in high-stakes testing grades lose the most time to test-taking and test preparation.

Our study used a detailed researched-based rubric for estimating instructional time lost to test preparation (narrowly defined as giving practice tests and teaching test-taking strategies, but *does not count* activities aligning content to the test such as review, reteaching or tutoring). The narrow definition of test prep yielded conservative estimates relative to other studies of test prep.

FINDINGS

Tests in the United States can be categorized into several groups, including:

- Large-scale standardized tests, usually summative, mandated by states primarily for accountability purposes;
- Interim/benchmarking tests used by districts to improve performance of low-achieving students on the state-mandated tests through targeted instructional interventions, review and reteaching, regrouping students, identifying students for tutoring and sharing results with students; and
- A variety of locally determined tests, including summative, diagnostic and miscellaneous tests, or formative assessment techniques by teachers.

Based on a detailed grade-by-grade analysis of the direct costs and the time costs of testing in the two school districts' assessment inventories, our study found:

- **Pervasive testing.** One of the districts in our study had 14 different assessments given to all students at least once a year in at least one grade. Some assessments are administered for several subjects multiple times a year resulting in 34 different test administrations. The other district's testing inventory had 12 different assessments but 47 separate administrations over the course of the year.
- **Test-taking time.** Students in one district in grades 3-10 spent approximately 15 hours or more per year (about three full school days) taking state-mandated tests, interim/benchmarking tests and other district academic assessments. Students in the other district in grades 6-11 devoted up to 55 hours per year to taking tests (about two full weeks of the school year).
- **Time for administrative tasks with students.** This includes giving directions, passing out test booklets and answer sheets, reading directions on the computer, etc., before and after each testing session. These administrative tasks with students took more than five hours annually—one full school day—in one of the districts. In the other district, administrative tasks with students used up more than 10 hours of the school year—two full school days—in the most highly tested grades.
- **Direct budgetary costs.** Several national studies show that the direct cost of purchasing, licensing and scoring state-mandated tests is around $25 per test-taker, and the annual cost of interim/benchmark testing is about $20 per test-taker. But when considering the cost of all the tests in a school district's inventory, the direct budgetary costs of the district testing program ranged from $50 per test-taker in one district to over $100 per test-taker in the other for grades 2-11. The direct budgetary cost of state testing represents less than 1 percent of K-12 per-pupil education spending. Nationally, education spending averages about $11,000 per pupil and reaches $20,000 per pupil in the highest-spending states.
- **Logistics and administrative costs.** Estimated at $2 per student per hour of testing (up to $80 per year for students in several grades in one district), these are costs associated with managing pallets of testing boxes; verifying and affixing data labels to test booklets, which could include three versions of the test at each grade level; and placing testing materials in secure locations before and after each round of testing to prevent cheating. After testing is completed, each school has to collect booklets, pack them and ship them off for scoring.
- **Test preparation time.** The detailed researched-based rubric narrowly defined "test preparation" to include giving practice tests and teaching test-taking strategies, but does not count review, reteaching or tutoring. Students in grades 3-8 in one district spent at least 80 hours per year (approximately 16 full school days) preparing for state-mandated tests, the

associated interim/benchmarking tests and all of the other district assessments. In the other district, students in grades 6-11 devoted 100 hours or more per year on test prep (approximately one full month of the school year).

- **The cost of testing and lost instructional time.** If school districts lengthen the school day or the school year to regain the instructional time lost to testing, the direct budget costs of testing are far from inconsequential. Adding one hour to the school day costs about $6.15 per student. In one district, the annual cost of testing per pupil in grades 3-8, including the cost of lost instructional time, was about $700—approximately 7 percent of per-pupil expenditures in the typical state. In the other district, the cost of testing in grades 6-11 exceeded $1,100 per student—about 11 percent of per-pupil expenditures in the typical state.

- **Alternate uses of testing time and costs.** Redirecting time and money devoted to testing to other uses would provide a lot more time for instruction—possibly including partial restoration of art, music and PE programs, during the existing school day. Cutting test prep and testing time in half could still restore significant minutes for instruction and would free up funding that could be used to purchase better tests, such as the Common Core assessments.

RECOMMENDATIONS

- **Testing for accountability.** Parents, teachers, principals, school districts, states and other stakeholders should be able to access the outcomes of standardized testing, but simple "snapshot" test and accountability measures should be no substitute for high-quality research and sophisticated evaluations of schools and educators.

- **Proper purpose of testing.** Tests should be used to improve instruction for students and to help improve schools—not to sanction, punish or close schools. Achievement tests should be just one of multiple measures used in a school evaluation system. Students should not take tests for the sole purpose of evaluating educators' performance.

- **State monitoring of the testing burden.** States and districts should conduct reviews of the cost of the entire inventory of assessments used in typical school districts, including test preparation time and lost instructional time.

- **Assessment literacy for educators.** Provide support to improve the assessment literacy of educators needed for developing and using classroom (formative) assessments and the proper use and interpretation of formal assessments, and then hold educators accountable for using best assessment practices. This includes teachers, as well as school, district and state administrators.

- **Streamline testing with teacher input.** School districts should review their entire assessment inventory *with teachers* to determine the usefulness of the assessments for improving instruction; unnecessary testing that doesn't support improved instruction should be eliminated.

- **Interim/benchmark testing.** Eliminate or sharply reduce this form of testing to no more than one administration per year. Most educators and many experts believe that interim/benchmark tests are not useful for improving instruction, and instead are ineffective low-quality tests, not aligned well to state standards presented in an overly elaborate, confusing manner; and used primarily to identify "bubble kids".

- **Common Core assessments.** Common Core states should adopt the "next generation" Common Core assessments designed by PARCC (Partnership for Assessment of Readiness for College and Careers) or SBAC (Smarter Balanced Assessment Consortium) as the only state-mandated assessments and eliminate all duplicative out-of-data state assessments, including the old recycled state tests claiming to be aligned to the Common Core.

- **Moratorium on high-stakes consequences.** Adopt a moratorium on the *high-stakes consequences* attached to all forms of state-mandated testing, including the Common Core assessments. Teachers need to reduce time spent on test prep and benchmark testing aligned to the old tests and instead focus their time on acquiring the skills and knowledge necessary to implement the Common Core standards in a threat-free environment.

- **Accelerate the development of embedded and rapid assessments.** Assessments embedded in instructional materials, using emerging technology and research on learning, have the potential to inform instruction better than time-consuming accountability tests through a balance of fine-grained classroom diagnostic tests, challenging tasks and projects in which students work through individual topics at their own pace, taking brief tests of their mastery along the way.

The Testing Landscape

State-mandated testing exploded in the 1990s when students in selected grades (usually one in elementary, one in middle school and one in high school) were tested once a year to measure school performance with respect to the nascent "standards" movement. Some research found that the first states to implement a standards-based reform with state-mandated testing improved student achievement faster than non-implementing states (Hanushek and Raymond, 2005). Teachers have always supported higher standards in large percentages, but they never found the end-of-year state tests useful for the specific purpose of improving instruction on a day-to-day basis, preferring instead the formative techniques and diagnostic tests.

The No Child Left Behind (NCLB) law, enacted in 2002, and subsequent federal mandates brought about a significant increase in accountability testing by requiring the testing of all students in grades 3-8 and some grades in high school. The federally mandated tests are used to sanction schools and, in extreme cases, to close some schools or convert them to charter schools. The federal Race to the Top (RTTT) program further fueled high-stakes testing by requiring teacher evaluations to be based in part on student achievement measured by test scores, as well as on other measures not related to testing. Subsequently, the federal government allowed states to receive waivers on some of NCLB's test-based sanctions if they implemented different test-based sanctions and enacted prescriptive teacher evaluation systems. The federal School Improvement Grant (SIG) program also threatens schools and staff with job loss, conversion to charter schools or closure based on student performance on state and federally mandated tests.

Schools use many types of assessments. Some are required by the state; others are chosen or developed at the district level. Some of a district's student assessment inventory is linked to federal laws and regulations other than NCLB. In many respects, given the ubiquitous emphasis on data reporting and analysis, nearly every assessment has become a high-stakes assessment with consequences for students, teachers and schools.

State-mandated summative tests are designed to assess whether students are meeting state standards in a given year. Tests are generally conducted in the spring as a means of "summing up" what students know that year, which is why these tests are referred to as summative assessments. The results of these tests are tied to the state's accountability system as mandated by NCLB. When linked to prior-year summative test scores, student progress can be measured, which provides some information about how well schools or teachers did in the current testing year, the true measure of school effectiveness. Unfortunately, many states and the federal government sanction schools for a single summative test score—a measure of student achievement but not school or teacher effectiveness—rather than the students' progress, which would show the school's contribution to learning over the course of the school year.

Both Midwestern School District (Table 1A) and Eastern School District (Table 1B) are implementing new state end-of-course testing in high school. NCLB requires testing in only one high school grade, but teacher evaluation requirements and tougher career- and college-readiness standards in the RTTT and NCLB waiver requirements have supported the push toward end-of-course assessments.

Eastern District students may take as many as 10 state-mandated end-of-course assessments (algebra 1, algebra 2, geometry, literature, English composition, biology, chemistry, U.S. history, world history, civics and government) by the time they finish high school. Midwestern District students may take as many as seven state-mandated end-of-course assessments (algebra 1, advanced algebra, geometry, English 9, English 10, American history, and government).

*Next-generation summative assessments***.** The PARCC- and SBAC-developed Common Core assessments, scheduled for full implementation in 2014-15, represent the next generation of summative tests. However, they still provide only a snapshot of student achievement at a specific point in time and will be used primarily for accountability purposes. Both the PARCC and the SBAC assessment consortia

Table 1A

Assessment Inventory, Midwestern School District
Minutes Per Test, Number of Annual Administrations and Total Annual Testing Minutes

Local Assessments	Grade		Minutes		Annual Minutes
DIAL	Pre-K	Developmental indicators	30	2x	60
NNAT2 – Naglieri Nonverbal	K,2,6,8	Ability	30	1x	30
DRA	K-6	Developmental reading	60-70	3x	180-210
Stanford Achievement Test	K	Reading, math	90	1x	90
	1	Reading, math	150	2x	300
	2	Reading, math	177	2x	354
21st Century Skills Test	5,8	Tech literacy	60	2x	120
	6-7	Tech literacy	60	1x	60
ACT EXPLORE	7-8	ACT prep	165	1x	165
ACT PLAN	9-10	ACT prep	175	1x	175
ACT	11	ACT prep	235	1x	235
NOCTI	12	Career/technical EOC	300	1x	300
Interim/Benchmark Tests					
ACUITY	3-6	Communication arts, math	180	3x	540
	7-8	Communication arts, math	180	1x	180
	5,8	Science	90	2x	180
Mock End-of-Course	9-12	Algebra, Eng., bio., govt.	50	2x	100
Scantron Performance Series	7-11	Communication arts, math	60	2x	120
State-Mandated Tests					
Grade 3-8 Assessments	3-8	Communication arts, math	270	1x	270
	5,8	Science	120	1x	120
End-of-Course Assessments	9-10	Communication arts	90	1x	90
	9-11	Math	90	1x	90
	9	Science	180	1x	180
	9,10	Social studies	90	1x	90

are developing computer-based testing systems that will be accompanied by benchmarking tools and use more innovative types of questions, online delivery, automated and artificial-intelligence powered scoring engines, and immediate Web-based reporting of results. Both Midwestern and Eastern school districts will fully implement the Common Core assessments in 2014-15.

Interim/Benchmarking Assessments. In addition to state-mandated summative testing, almost all districts now have some type of interim or benchmarking assessments, which are essentially duplicative summative assessments, to measure student growth toward passing the state-mandated test over the course of a school year (Heppen et al., 2011). Unlike the state-mandated summative assessments, results from these assessments are available during the school year and are believed to help improve scores on state-mandated tests when used to target instructional interventions, to review and reteach, to regroup students, to identify students for tutoring, and to share results with students. A survey of curriculum coordinators and research directors revealed

Table 1B

Assessment Inventory, Midwestern School District
Minutes Per Test, Number of Annual Administrations and Total Annual Testing Minutes

Local Assessments	Grade		Minutes		Annual Minutes
Kindergarten Test	K	Developmental indicators	30	3x	90
DIBELS	K-3	Reading diagnostic	10	3x	30
	4-5		10	1x	10
SRI (Scholastic)	6-9	Reading	30	3x	90
TerraNova	K-1	Reading, math	120	1x	120
	2		130	1x	130
Curriculum-Based Assessments (CBA)	2-3	Math	50	3x	150
	6-12	Math	50	2x	100
		Reading	75	2x	150
		Science	110	2x	220
		Social studies	110	3x	330
PSAT	10,11	Preliminary SAT	135	1x	135
NOCTI	12	Career/technical EOC	300	1x	300
GRADE (Pearson)	3-11	Literacy (federal grant)	90	3x	270
Interim/Benchmark Tests					
Curriculum-Based Assessments (CBA)	3-5	Math	50	4x	200
Classroom Diagnostic (CDT)	6-11	Reading	75	5x	375
		Math, reading, science	270	3x	810
State-Mandated Tests					
Grades 3-8, 11	3-8	Math, reading	395	1x	395
	4, 8	Science	160	1x	160
	5, 8	Writing	330	1x	330
	11	Math, reading, writing, science	990	1x	990
End-of-Course*	9-12	Math, reading, social studies, science	120	2x	240

*Assumes students take two end-of-course assessments per year for four years.

that while nearly all respondents believed that the purpose of interim assessments was to guide and improve instruction, 90 percent of the same respondents believed they measure progress toward the end-of-year state test, 80 percent believed they were also diagnostic and 90 percent thought they were formative (Heppen et al., 2011).

Despite the ubiquity of interim/benchmark testing, research on these tests has failed to demonstrate a positive and statistically significant impact on improving student achievement on end-of-year tests, which is the tests sole purpose.[1] The studies that have found some positive impacts focused more on data use and accessibility."[2] It is not clear whether the

1. A national study conducted by Mathematica Policy Research (Furgeson et al., 2012) of 22 Charter Management Organizations (CMOs) found that "frequent formative student assessments" (referring to interim/benchmarking assessments) had no impact on positive achievement impacts among CMOs. In a study of Formative Assessments of Student Thinking in Reading (FAST-R), teachers in Boston were given test data aligned to the state assessments every three to 12 weeks, and data coaches helped teachers interpret and use the data. A two-year evaluation of 21 elementary schools found small but statistically insignificant effects (Quint et al., 2008). A one-year study of benchmark assessments in 22 Massachusetts middle schools also showed no effect (Henderson et al., 2007). May and Robinson (2007) studied a benchmark assessment program used in high schools to prepare students for the Ohio Graduation Tests and found no statistically significant impact from the benchmarking for students taking the graduation test for the first time.

results of those studies are due to the benchmarking tests, the practice effects of repeated test-taking, or the improved data literacy of teachers. New Leaders for New Schools (Marshall et al., 2006) identified several reasons why interim assessments may not produce significant achievement gains:

- Poor alignment with state standards, tests or pacing guides;
- Results presented in an overly elaborate, confusing manner;
- Results used only to focus on bubble kids;
- Not administered frequently enough to have an impact on instruction;
- Scored externally with teachers having no investment in the results;
- Too short and don't give teachers enough detailed data; and
- Teachers think tests will be used to blame them.

Both Midwestern School District and Eastern School District layered on an array of interim/benchmarking tests administered several times a year:

Midwestern School District: ACUITY is a state-tailored commercial benchmarking test administered three times annually. The Scantron Performance Series, given twice annually, is also used for benchmarking in grades 7-11. Midwestern District has "mock" end-of-course testing twice a year in high school.

Eastern School District: The state-developed Classroom Diagnostic Tool (CDT), administered three times annually, is used to predict outcomes on the state-mandated test in grades 6-11 and is also used for benchmarking to the state end-of-course tests. For benchmarking purposes in grades 3-5, the district used a modified version of the district-developed Curriculum-Based Assessments, which is administered four times a year in math and five times a year in reading.

Classroom-based assessments. These assessments are used during teaching and are embedded in instruction; they are a tool that helps teachers adjust their instruction in the moment to meet the needs of students. This isn't just about giving a quiz at the end of class, but it's also questioning and observing students on performance tasks. Results are received instantly, which allows teachers to adjust their instruction immediately. Teachers also use tests, quizzes and homework to assess student learning. Teachers surveyed in Chicago spent 22 minutes a day giving students curriculum subject assessments (tests, quizzes, etc.) and another 32 minutes a day assessing students' work during contractual hours (Bruno, Ashby and Manzo, 2012).

Embedded and rapid assessments. Embedded assessments could be thought of as the "next generation" formative assessment. Assessments embedded in instructional materials using emerging technology and research on learning aims to inform instruction through a balance of fine-grained classroom diagnostic tests, challenging tasks and projects in which students work through individual topics at their own pace, taking brief tests of their mastery along the way, with feedback delivered to the student and teacher on individual processes or misconceptions that cause the student problems (Gordon Commission, 2013). They are now used in some computer-based math programs such as Carnegie Learning, Khan Academy and Agile Mind, and in Scholastic's READ 180 and Lexia Reading. Both Eastern and Midwestern school districts already use READ 180, and Eastern School District also uses a Carnegie Learning program.

Randomized and quasi-experimental research indicate that "rapid formative assessment systems" effectively raise student achievement (Nunnery et al., 2006; Ross et al., 2004; Ysseldyke and Bolt, 2007; Ysseldyke and Tardrew, 2007) and are a more efficient use of resources than a range of alternative interventions such as a longer school day, value-

2. A multidistrict, multistate quasi-experimental study of the first year of a data-driven reform strategy (Carlson, Borman, Robinson, 2011) found a small positive impact on math scores but no statistically significant impact on reading scores. However, subsequent research on all four years of the same program (Slavin et al., 2013) revealed that impacts on elementary reading and math were never statistically significant in any of the four years except fourth-year reading. In middle school, significant positive effects were limited to the first two years in reading and only the first year in math. In a value-added study of four districts (102 principals and 593 teachers), an index of teacher general data use had a positive impact on middle-grades math and elementary-grades reading, but no impact on middle school reading or elementary math (Faria et al., 2012). An index of principal general data use had a positive impact on middle-grades math but no impact on middle-grades reading or at the elementary level for either subject.

Midwestern Public Schools

Testing Calendar

Aug./Sept.

SUN	MON	TUE	WED	THU	FRI	SAT
18	19	20	21	22	23	24
25	26	27	28	29	30	31
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

February

SUN	MON	TUE	WED	THU	FRI	SAT
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	1

October

SUN	MON	TUE	WED	THU	FRI	SAT
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

March

SUN	MON	TUE	WED	THU	FRI	SAT
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

November

SUN	MON	TUE	WED	THU	FRI	SAT
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

April

SUN	MON	TUE	WED	THU	FRI	SAT
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

December

SUN	MON	TUE	WED	THU	FRI	SAT
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

May

SUN	MON	TUE	WED	THU	FRI	SAT
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

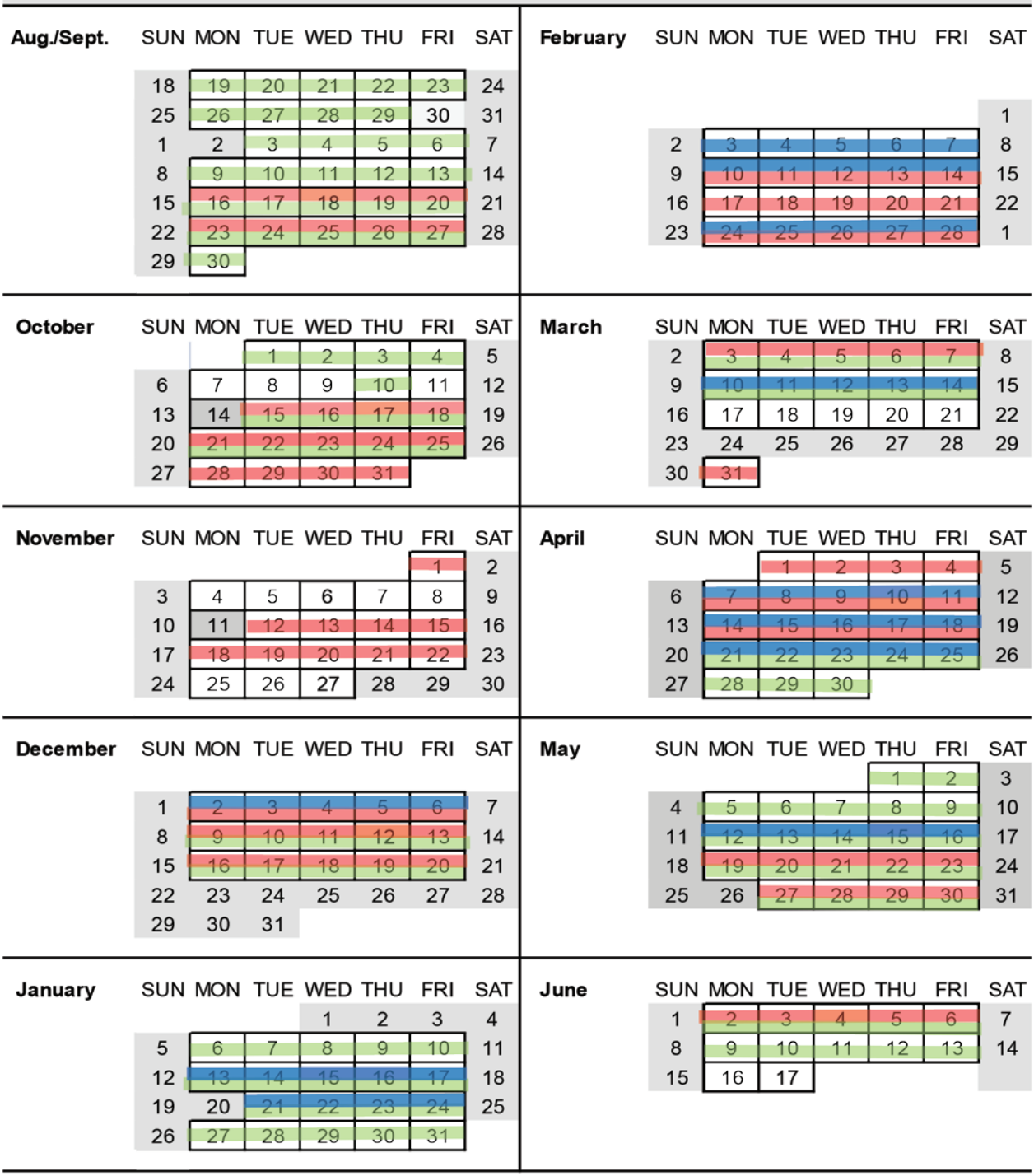
January

SUN	MON	TUE	WED	THU	FRI	SAT
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

State-mandated testing window for at least one grade and subject.
Interim/benchmarking testing window for at least one grade and subject.
Local testing window for at least one assessment, grade and subject.

Eastern School District

Testing Calendar



State-mandated testing window for at least one grade and subject.
Interim/benchmarking testing window for at least one grade and subject.
Local testing window for at least one assessment, grade and subject.

added teacher assessment, class size reduction, or full-day kindergarten (Yeh, 2008; Yeh, 2010). The rapid assessment systems in these studies have students choose books based on their reading levels and then they take a comprehension quiz after reading each book. In math, each student receives individualized math problem sets daily, and students score their answers using a device in the classroom. The rapid assessment system imposes a minimal burden on teachers because students scan their own bubble sheets, software is used to score each assessment, and reports are available to teachers and administrators electronically.

Diagnostic Testing. Districts use various assessments to ascertain each student's strengths and weaknesses to adjust curricula and instruction to meet student needs. Some states require districts to administer assessments to students on basic literacy. Some diagnostic assessments are given to all students in the early grades but only to low-scoring students in subsequent grades.

Midwestern School District: The DIAL assessment of developmental indicators is given to all pre-K students; the Naglieri nonverbal ability test is administered in kindergarten and grades 2, 6 and 8. The Developmental Reading Assessment (DRA) is administered three times annually in kindergarten through grade 6. Both the DIAL and the DRA are administered and scored by teachers. Although strictly speaking not a diagnostic test, the Stanford Achievement Test is given once a year to kindergartners and twice annually to students in grades 1-2—those grades not subject to mandated state testing.

Eastern School District: A kindergarten assessment of developmental indicators is individually administered to each student three times annually followed by the DIBELS language assessment, also administered three times a year, to all students in grades K-3 and once annually in grades 4-5. Technically not a diagnostic test, the TerraNova achievement test is given annually in grades K-2.

College Entrance Exams and Advanced Placement. College-bound students are often interested in taking Advanced Placement or International Baccalaureate courses, and this course work has national tests. Midwestern District has a multiyear program to prepare students for the college entrance examination. The EXPLORE (Grades 7-8), PLAN (Grades 9-10) and ACT District Choice Testing are pre-ACT tests that project ACT scores as well as data regarding concepts and skills that students can improve to help progressively increase their performance. Most students in the district participate in PLAN and EXPLORE. In Eastern District, students in grades 10 and 11 take the Preliminary SAT (PSAT).

Other Academic Assessments. Districts may choose, at their discretion, to administer other assessments, although these locally chosen assessments may be related to a federal or state mandate. Midwestern School District, for example, adopted the 21st Century Skills Assessment for its NCLB Technology Literacy measure. This test is given to students in grades 5 and 8 twice a year; in grades 6-7 it is given once in the spring.

Eastern School District has implemented an extensive district-developed Curriculum-Based Assessments (CBA) system, which is administered at least twice yearly. The CBAs are "standardized" across the district in grades 2-12, but they are teacher-developed and teacher-scored. The CBAs overlap with state-mandated testing, and in grades 3-5 modified versions of the CBAs (administered four times a year in math and five times a year in reading) are used as interim/benchmarking tests. The GRADE (Pearson) literacy assessment adds a fourth layer of tests (in addition to state tests, benchmarking and the CBAs). The 90-minute test, which is given three times a year, is a requirement for participating in a federally funded literacy program.

Time Committed to Testing

Time for Students to Read and Respond to Questions. Most students in Midwestern School District spent approximately 15 hours or more per year on state-mandated tests, interim/benchmarking tests and other district assessments (Figure 1A). The time committed to testing in Eastern School District—where students in grades 6-11 spent up to 55 hours per year on state-mandated, benchmarking tests and district assessments (Figure 1B)—was substantially higher than in Midwestern School District.

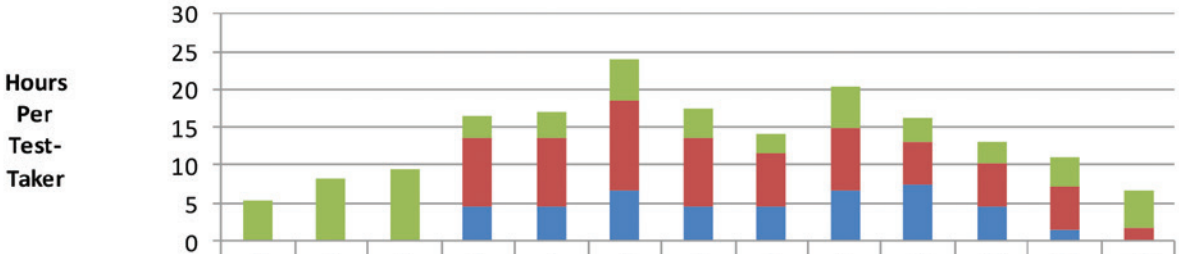
Midwestern School District: Fifth- and eighth-grade students spent the most time on tests (more than 20 hours per pupil); students in K-2 and grades 11-12 spent the least time (five to 10 hours per pupil). The state-mandated tests exceeded five hours only in the fifth, eighth and ninth grades, in part because the state tests use only multiple-choice questions. Many states add short-answer or extended-response questions that demonstrate performance but require costly grading by hand. The interim/benchmarking tests are more time-consuming than the state-mandated tests.

Eastern School District: Eighth- and 11th-grade students spent the most time on tests (50 to 55 hours per year) and K-2 students the least (about five hours per year). The district tests are the most time-consuming of all of the tests examined, amounting for more than 30 hours per year in grades 6-11. Interim/benchmarking tests exceeded nine hours per year in grades 3-5 and exceeded 13 hours per year in grades 6-11.

Time for Administrative Tasks with Students. In addition to the time students are given to read test items and answer questions, students sit in testing sites while administrative tasks are being performed (e.g., receiving directions, as test booklets and answer sheets are passed out, reading directions on the computer, etc.) before and after each testing session. The widely used MAP interim/benchmark test, for example, allows about an hour per subject for students to read and answer questions, but MAP wants districts to reserve 75 minutes to allow for administrative tasks with students (NWEA, 2006).

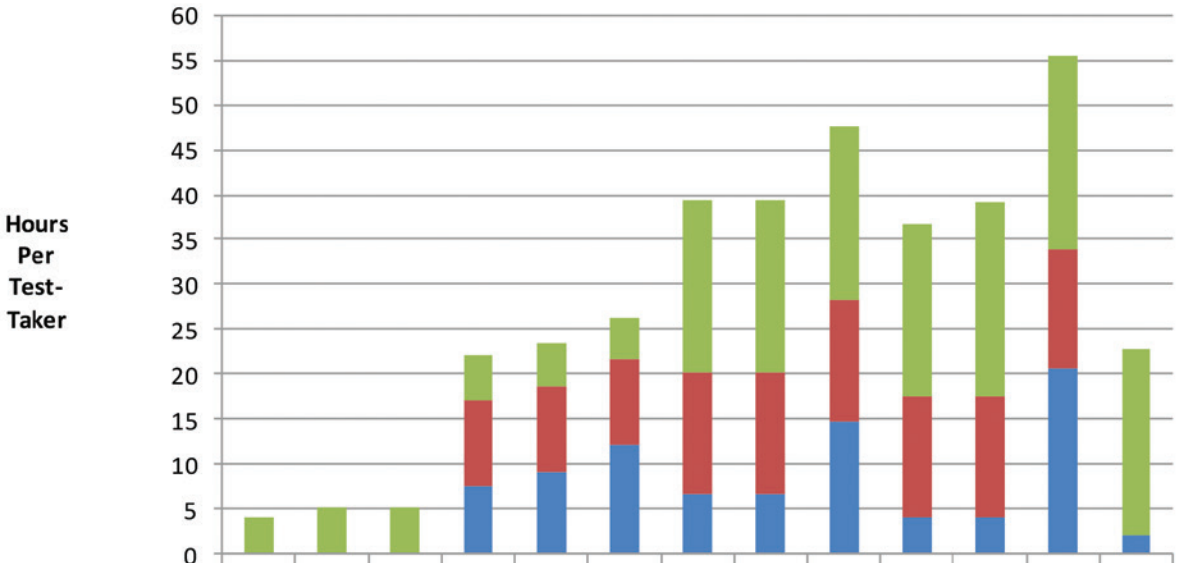
In Pennsylvania, state officials recommend that district test administrators add 20 minutes for each testing session to account for administrative tasks with students (Pennsylvania Department of Education, 2011). For example, the three math testing sessions add up to 185 minutes, but test administrators need to schedule another 60 minutes (three sessions at 20 minutes each) for administrative tasks with students. Over all four subjects, the 12 testing sessions add

Figure 1A - Testing Time Per Student Per Year, Midwestern District



	K	1	2	3	4	5	6	7	8	9	10	11	12
Local	5.3	8.1	9.4	3.0	3.5	5.5	4.0	2.8	5.3	2.9	2.9	3.9	5.0
Benchmark	0.0	0.0	0.0	9.0	9.0	12.0	9.0	7.0	8.5	5.7	5.7	5.7	1.7
State-Mandated	0.0	0.0	0.0	4.5	4.5	6.5	4.5	4.5	6.5	7.5	4.5	1.5	0.0

Figure 1B - Testing Time Per Student Per Year, Eastern District



	K	1	2	3	4	5	6	7	8	9	10	11	12
Local	4.0	5.0	5.2	5.0	4.7	4.7	19.3	19.3	19.3	19.3	21.6	21.6	20.8
Benchmark	0.0	0.0	0.0	9.6	9.6	9.6	13.5	13.5	13.5	13.5	13.5	13.5	0.0
State-Mandated	0.0	0.0	0.0	7.6	9.1	12.1	6.6	6.6	14.8	4.0	4.0	20.5	2.0

approximately four hours to student testing time to account for administrative tasks with students—nearly a full day of instruction out of each school year.

Online testing may require more time for administrative tasks with students even though test booklets, pencils and scoring sheets are not used. Pennsylvania's interim/benchmarking test, the Classroom Diagnostic Tool (CDT), is an online computer adaptive test. Before taking the CDT, students take online assessment tutorials (nine to 15 minutes) for each course tested, and they are allowed to repeat the tutorial as often as desired or needed. Pennsylvania recommends giving students at least 20 minutes to go through all questions in the online tools training for each course tested.

For estimates of time spent on testing, our study adds 20 minutes per testing session for administrative tasks with students. Table 2 shows the hours per grade consumed for administrative tasks with students that need to be added to the time students spend reading and answering test questions. In Eastern School District, students in many grades spend two full school days just on getting directions, handing out test booklets, handing in answer sheets and similar administrative activities.

Table 2

Time for Administrative Tasks with Students State-Mandated, Benchmarking and Local Tests

	Midwestern District		Eastern District	
	Sessions	Hours	Sessions	Hours
K	6	2.0	2	0.7
1	11	3.7	5	1.7
2	12	4.0	5	1.7
3	15	5.0	18	6.0
4	15	5.0	20	6.7
5	21	7.0	22	7.3
6	16	5.3	29	9.7
7	17	5.7	29	9.7
8	23	7.7	35	11.7
9	16	5.3	25	8.3
10	14	4.7	27	9.0
11	12	4.0	41	13.7
12	7	2.3	19	6.3

Direct Cost of Tests

Costs of Test Development and Implementation. A recent Brookings Institution study (Chingos, 2012) calculated that states spent roughly $27 annually per student tested on contracts for state-mandated tests. In another highly regarded study (Topol et al., 2012), the Assessments Solutions Group (ASG) calculated higher state costs than the Brookings study. Finding that the average current state spending on mathematics and ELA summative assessments required by NCLB is approximately $25 per student tested, additional state spending on non-NCLB required assessments (additional assessments, subjects and grades) added roughly $10 per pupil tested. In addition, the ASG study concluded that school districts are spending an average of $15 to $20 or more per student on interim/benchmarking assessments. The combined costs of typical state and local spending on interim and summative assessments in ELA and mathematics are in the vicinity of $35 to $55 per pupil tested.

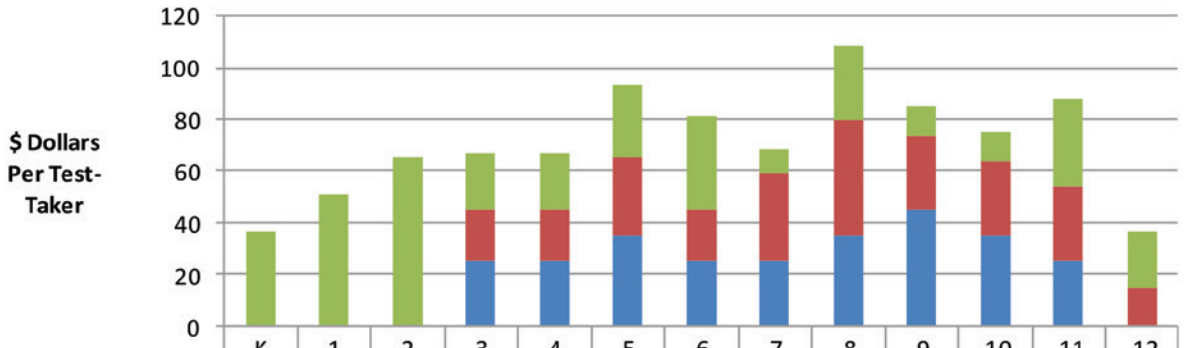
Our study's calculation of direct costs of testing applies the ASG direct cost estimates for state-mandated tests ($25 for ELA and math plus $10 for each additional test) and interim-benchmarking tests ($20) to the testing regime in Midwestern and Eastern school districts. Additionally, our study adds the direct costs of other local assessments based on costs reported by the school districts or on prices charged to school districts by the test developers on the Internet. Appendix Tables A and B describe the grade-by-grade, test-by-test direct cost estimates.

Midwestern School District spent more than $80 per pupil tested in grades 5-6 and 8, but less than $60 per pupil in grades K-1 and 12. See Figure 2A.

Eastern School District spent around $50 to $70 per pupil tested in most grades. Grade 8 costs hit $80 per pupil, and grade 11 costs surpassed $100 per pupil tested. See Figure 2B. CBAs are not included as a direct cost in Figure 2B because they are developed by district teachers and staff rather than purchased from a vendor.

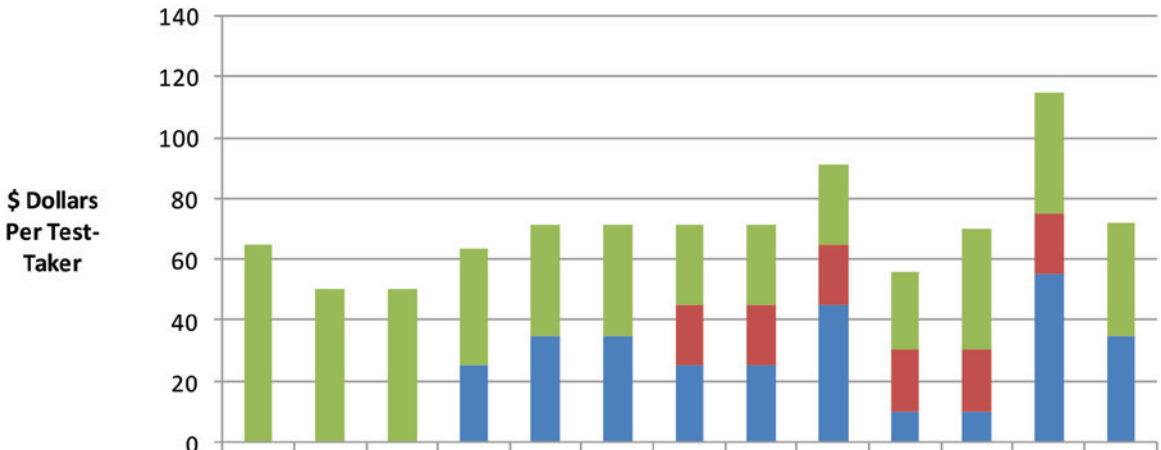
The direct cost of state testing in the two districts represents less than 1 percent of per-pupil spending on K-12 education. U.S. education spending averages about $11,000 per pupil in 2010-11 and reaches

Figure 2A - Direct Costs of Assessments and Licensing Per Test-Taker, Midwestern District



	K	1	2	3	4	5	6	7	8	9	10	11	12
Local	$36	$51	$65	$22	$22	$28	$36	$9	$29	$11	$11	$34	$22
Benchmark	0	0	0	$20	$20	$30	$20	$34	$44	$29	$29	$29	$14
State-Mandated	0	0	0	$25	$25	$35	$25	$25	$35	$45	$35	$25	0

Figure 2B - Direct Costs of Assessments and Licensing Per Test-Taker, Eastern District



	K	1	2	3	4	5	6	7	8	9	10	11	12
Local	$65	$50	$50	$39	$36	$36	$26	$26	$26	$26	$40	$40	$37
Benchmark	0	0	0	0	0	0	$20	$20	$20	$20	$20	$20	0
State-Mandated	0	0	0	$25	$35	$35	$25	$25	$45	$10	$10	$55	$35

$20,000 per pupil in the highest-spending states (Hussar and Bailey, 2011; Cornman et al., 2013).

Costs of Logistics and Administrative Support. In a study of Wisconsin's state-mandated testing program, Zellner, Frontier and Pheifer (2006) describe the cost of logistics and administrative support as, "a broad range of efforts from school staff. Pallets of testing boxes arrive in districts. Before testing, data labels generated by the state must be verified for accuracy and affixed to individual test booklets. Booklets, which include three versions of the test at each grade level, must be placed in secure locations before and after each round of testing. After testing is completed, each school has to collect booklets, pack them, and ship them off for scoring."

Table 3 uses the time data from the Wisconsin survey to reach an estimate of the costs of logistics and administrative support for testing:

The average school district in the Wisconsin survey enrolled 1,432 students. The state-mandated test applied to students in grades 3-8 and 10, or approximately 800 students. Overall, the cost of logistics and administration averaged $14 per student tested and was composed of the following elements:

- *Paraprofessionals* spent a per-district average of 102 hours engaged in duties ranging from giving tests to small groups of test-takers to assisting teachers with whole-class testing.
- *Administrators* spent a per district average of 62 hours engaged in a variety of testing-related tasks. Some schools had to modify schedules and readjust staffing needs for the several days needed for testing.
- Test administration required *substitute teachers* to proctor tests or supervise the classrooms of teachers who were engaged in other testing tasks. Across all districts in the sample, 1,021 substitute teachers facilitated testing or supervised students.
- *Guidance counselors* spent 92 hours, a greater percentage of their time facilitating the testing process than any other group, suggesting a marked loss of guidance services.

Wisconsin students spent between 4.75 to 8.66 hours a year on the state-mandated tests. The $14

Table 3

Logistics and Administrative support
Average District Cost

	Total Hours	Hourly Cost	Total Cost	Pupils Tested	Cost Per pupil
Secretaries	91	$23	$2,093	800	$2.62
Counselors/test coordinators	92	$40	$3,680	800	$4.57
Paraprofessionals	102	$13	$1,326	800	$1.68
Administrators	62	$51	$3,162	800	$3.96
Substitute teachers	48	$20	$960	800	$1.20
					$14.03

per student cost for approximately 7 hours of testing yields a logistics and administration cost estimate of $2 per student, per hour of testing.

The logistics and test administration burden is likely to be comparable, if not higher, for tests taken on computers. The computer-delivered MAP interim/benchmarking test (NWEA, 2012), for example, lists numerous administrative activities needed before the test:

- Train new proctors, new teachers and staff.
- Ensure technical readiness of computer labs.
- Have the latest version of the test-taking software installed.
- Create and submit class roster files two weeks before testing begins.
- Download student data and test packages.
- Distribute proctor and teacher usernames and passwords.

During the testing window, test results need to be updated daily. After testing, end-of-term reports need to be ordered, downloaded and/or printed.

Instructional Time Lost to Test Preparation

Time spent on test preparation reduces instructional time by the same amount, giving rise to great concern among educators and parents. Test-preparation time has grown substantially with the expansion of state-mandated testing, the proliferation of interim/benchmarking assessments and the expansion of high-stakes consequences attached to test scores.

Rubric for Measuring Classroom Time spent on Test Preparation. Our study estimates the amount of time teachers spend with students on test preparation using the rubric in Table 4, compiled mainly from two studies of state-mandated testing described in the Appendix—one by the Center for the Study of Testing, Evaluation, and Educational Policy (Pedulla et al., 2003) and one by the Consortium on Chicago School Research (Tepper Jacob and Stone, 2005).

The research on which the rubric is based narrowly defined test prep as giving practice tests and teaching test-taking strategies but does *not count* activities aligning content to the test such as review, reteaching or tutoring. The research shows that teachers prepared their students for testing even in the absence of high-stakes accountability, but that high-stakes considerably increased test-prep time. Further, our study assumes that teachers prepare students for the interim/benchmarking assessments as if they were low-stakes tests rather than treating them as part of the practice testing for the high-stakes assessments. The test-prep rubric is based on studies that predate interim/benchmark testing. As with the other low-stakes tests, and unlike test prep, the interim/benchmarking tests are part of district data systems and used to allocate instructional time and resources to improve state-mandated testing outcomes.

The rubric assumes at least two hours of classroom test prep for any kind of test except for individually administered one teacher-one student diagnostic tests, such as DIBELS. In low-stakes situations in both test-prep studies reviewed, teachers reported spending about 10 hours annually on test prep for a test given once a year covering two or more subjects. The rubric assumes five hours for a one-subject test given once a year.

High-stakes for students or schools increased test prep to 30 hours per year for elementary teachers in high-stakes grades and to 20 hours per year for high school teachers in high-stakes classrooms. The Center for the Study of Testing, Evaluation, and Educational Policy (Pedulla et al., 2003) reported high school teachers spending less time on test prep. The rubrics in Table 4 apply to average schools but the Consortium on Chicago School Research (Tepper Jacob and Stone, 2005) found that teachers in low-performing schools with high percentages of minority students reported more than 35 hours of test prep.

The rubric applies to each test administration regardless of the length of the test, but test-prep time estimates vary with the number of subjects tested. When a test was given multiple times a year, our study assumes that the first administration of the test involved minimal preparation (two hours) because there are no consequences attached to poor

Table 4

Rubric for Estimating Classroom Test-Preparation Time

	Hours Per Test Administration
Diagnostic tests (individually administered)	0
Diagnostic tests (group administered)	2
Pretests (first test, if test given multiple times a year)	2
Low-stakes test, low-stakes grades (one subject)	5
Low-stakes test, low-stakes grades (two subjects)	10
Low-stakes test, high-stakes grades (elementary)	15
Low-stakes test, high-stakes grades (high school)	10
High-stakes tests, high-stakes grades (elementary)	30
High-stakes tests, high-stakes grades (high school)	20

baseline student performance (and perhaps even motivation for poor student performance in order to inflate test score gains in subsequent test administrations). Interim/benchmark tests were treated as low-stakes tests but were almost always given in high-stakes grades, which resulted in the assumption of 15 hours of test prep for each administration after the first for elementary teachers.

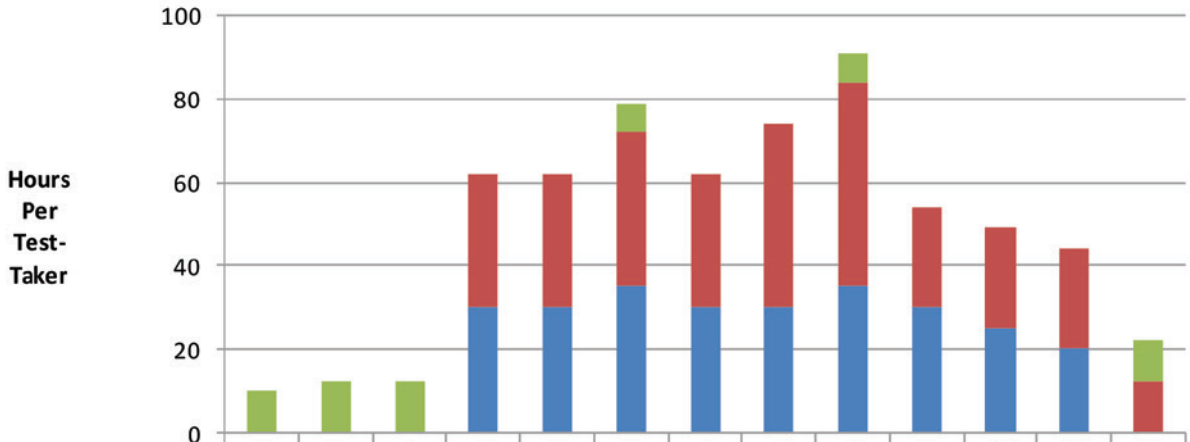
Estimates of Time Spent on Test Preparation. Students in grades 3-8 in Midwestern District spent at least 60 hours per year preparing for state-mandated, interim/benchmarking and district assessments (Figure 3A). The time committed to testing in Eastern District was substantially higher than in Midwestern District, with students in grades 6 -10 spending approximately 110 hours or more per year on test prep across all assessments (Figure 3B).

Midwestern School District: Fifth- and eighth-grade students spent the most time on test prep (more than 90 hours per pupil, or 18 full days, in eighth grade) while students in grades K-2 had 10-12 hours of test prep per year. Except for seniors, high school students spent around 45 to 55 hours in test prep per year. Estimates for test prep for the state-mandated tests never exceeded 35 hours per year. However, the test prep for the interim/benchmarking tests amounted to more than the test prep for the state-mandated tests in grades 3-8. Appendix Table C provides more detailed information about the hours spent on test prep by test and grade.

Eastern School District: In grades 3-5, the local CBAs have been developed into an interim/benchmarking system (administered five times annually in reading and four times annually in math). Test prep escalates after grade 5 because the district uses both the state-prepared benchmarking tests and district-developed CBAs. In grades 6-12, the CBAs are listed as local tests. Test prep for interim/benchmarking tests consumed about the same number of hours as test prep for state-mandated tests in grades 3-7 and slightly less time in grades 8-11. See Appendix Table D for estimates of test prep time by test and grade.

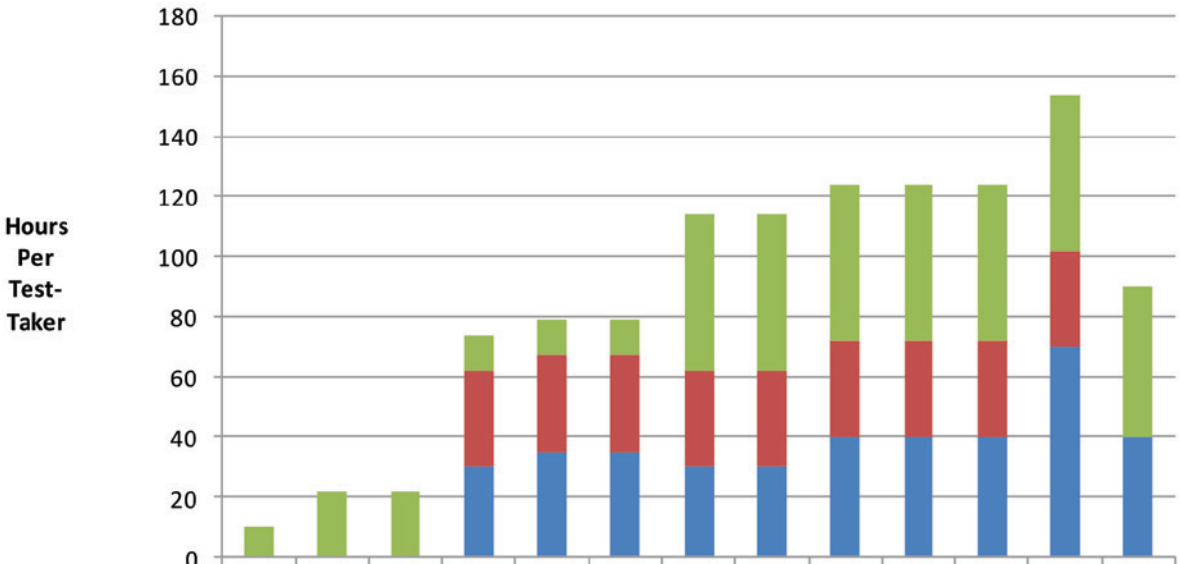
Our test-prep time estimates for grades 3-8 in Midwestern School District amounted to approximately two hours per week. In grades 6-11 in Eastern School District, test prep time totaled ap-

Figure 3A - Estimated Annual Test Preparation Time, Midwestern District



	K	1	2	3	4	5	6	7	8	9	10	11	12
Local	10	12	12	0	0	7	0	0	7	0	0	0	10
Benchmark	0	0	0	32	32	37	32	44	49	24	24	24	12
State-Mandated	0	0	0	30	30	35	30	30	35	30	25	20	0

Figure 3B - Estimated Annual Test Preparation Time, Eastern District



	K	1	2	3	4	5	6	7	8	9	10	11	12
Local	10	22	22	12	12	12	52	52	52	52	52	52	50
Benchmark	0	0	0	32	32	32	32	32	32	32	32	32	0
State-Mandated	0	0	0	30	35	35	30	30	40	40	40	70	40

proximately three hours per week. These estimates are low when compared with many other studies. In North Carolina, 80 percent of teachers in one study indicated that "they spent more than 20 percent of their total instructional time practicing for the end-of-grade tests" (Jones et al., 1999). Teachers in New York City report approximately the same percentage as the North Carolina survey. More than half of the teachers in New York said they spent five hours a week on test prep, the equivalent of one full day of instruction each week (Lucadamo, 2005). Likewise, in a Chicago teacher survey (Bruno, Ashby and Manzo, 2012), the average amount of additional time consumed monthly by prepping and administering mandated standardized tests was 20 hours (approximately one a day, or five hours per week).

But our very conservative findings of two to three hours of test prep per week reflect our narrow definition of test prep as giving practice tests and teaching test-taking strategies but *not counting* activities aligning content to the test such as review, reteaching or tutoring. In Texas, for example, the 89 percent of teachers who said, "I have been encouraged to conduct drill and practice on the TAKS objectives where I work" also reported that on average they spent two hours a week conducting "drill and practice on the TAKS objectives where I work" (Texas Federation of Teachers, 2005). But the same teachers also reported using an average of 40 percent of class time "devoted to preparation for the TAKS."

Total Cost of Testing

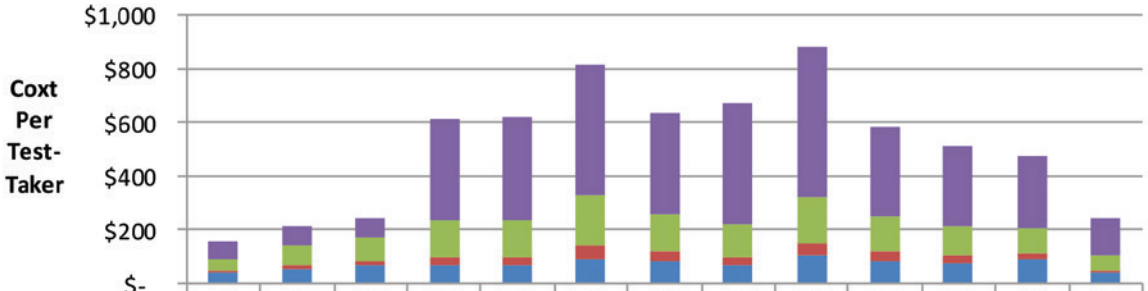
Expanding instructional time is an important reform. The time for students to take tests and prepare for tests is not a budgetary cost, but it is paid for through the reduction of instructional time in the exact amount by which testing time increases. The evidence of instructional time lost to test-taking and test prep suggests that resources for expanded learning time should concentrate on upper elementary school and middle school where testing consumes the most instructional time.

Expanded learning time is often needed just to compensate for expanded testing. The Hamilton Project at the Brookings Institution (Fryer, 2012) is an example of a reform agenda asking for both more learning time and more time for taking tests and using data. The report identified five educational practices that are most successful for improving student achievement; two of them were "using student data to drive instruction" and "increasing the length of the school day and the number of days in the school year." Data-driven instruction was defined as conducting regular assessments of students every four to six weeks, more in-depth assessments given several times a year, and teacher meetings with individual students to discuss and set goals for the state test after each interim assessment.

In addition to reducing instructional time to make room for ever increasing test-taking and test prep, high-stakes assessments may distort the educational process by shifting instructional time into subjects covered on the high-stakes exams and away from other core subjects like science and social studies, or from enrichment activities such as art and music.

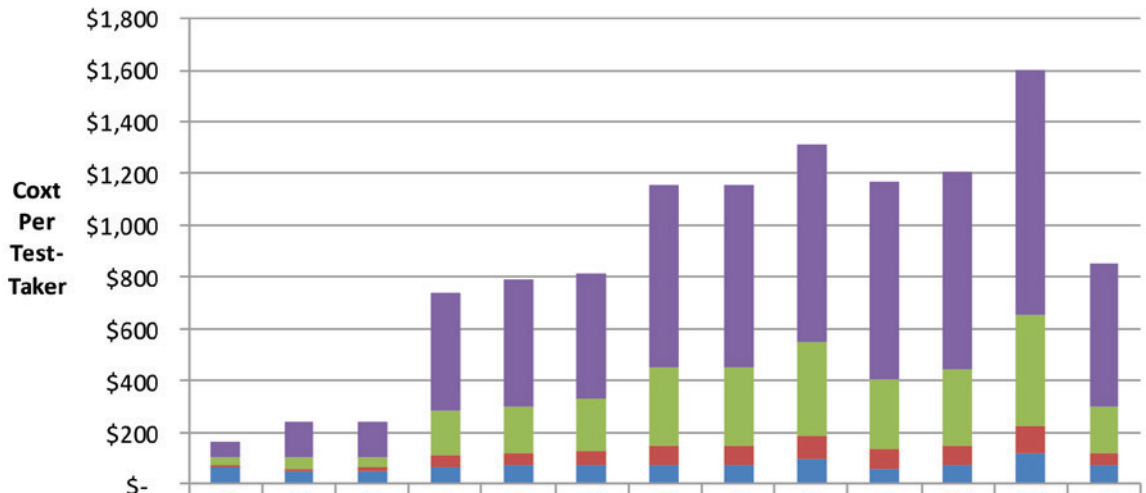
Estimated Cost of Instructional Time Lost to Student Testing. Increasing learning time to compensate for test-taking and test prep time is expensive. Experience with expanded learning time demonstrates that costs do not rise in proportion to the increase in time (Roza and Hawley, 2008). School-related, non-teaching costs such as facilities, transportation and administration do not rise automatically with the addition of time. Second, not all teachers participate in extended time. According to Education Sector (Silva, 2007), several analysts have suggested that a 10 percent increase in school time would increase education operating costs by 6 to 7 percent. The Massachusetts Expanded Learning

Figure 4A - Annual Cost of Testing and Lost Instruction Time
Midwestern District



	K	1	2	3	4	5	6	7	8	9	10	11	12
Test Prep	$62	$74	$74	$381	$381	$486	$381	$455	$560	$332	$301	$271	$135
Test Taking	$45	$72	$83	$132	$135	$191	$140	$122	$172	$132	$109	$93	$55
Logistics	$11	$16	$19	$33	$34	$48	$35	$29	$41	$32	$26	$22	$13
Direct Cost	$36	$51	$65	$67	$67	$93	$81	$68	$108	$85	$75	$88	$36

Figure 4B - Annual Cost of Testing and Lost Instruction Time
Eastern District



	K	1	2	3	4	5	6	7	8	9	10	11	12
Test Prep	$62	$135	$135	$455	$486	$486	$701	$701	$763	$763	$763	$947	$554
Test Taking	$29	$41	$42	$173	$185	$207	$302	$302	$364	$278	$296	$426	$179
Logistics	$8	$10	$10	$44	$47	$53	$79	$79	$95	$74	$78	$111	$46
Direct Cost	$65	$50	$50	$64	$71	$71	$71	$71	$91	$56	$70	$115	$72

Time Initiative estimated that increasing time by 30 percent (two hours per day) would cost an additional 20 percent of average operating expenditures (Silva, 2007).

A study by the Center for American Progress (Roza and Hawley Miles, 2008) provides a model for estimating the cost of expanding learning time by approximately 30 percent (equaling two hours per day or 360 hours per year on top of a 6.5 hour day over a 180-day instructional school year). The national average operating expenditure for 2012-13 is projected at $11,081 per student according to *Projections of Education Statistics to 2020* (Hussar and Bailey, 2011). Based on the assumption that increasing instructional time by 30 percent (two hours per day) raises expenditures by 20 percent, costs would increase by $2,216 per student for the 360 additional hours (or $6.15 per hour, per student).

The annual cost of testing for Midwestern School District (Figure 4A) and for Eastern School District (Figure 4B) combines: (1) direct costs for test purchasing and licensing; (2) costs for logistics and administration; and (3) time costs at $6.15 per hour, per student tested to account for instructional time lost to test-taking, test administration time with students, and test preparation.

Midwestern School District: The annual cost per pupil of student assessments in grades 3-8, including the cost of lost instructional time, is at least $600 per student. In grades K-2, testing costs are around $200 per student. In high school, except grade 12, per-student testing costs are in the $400 to $600 range.

Eastern School District: The annual cost per pupil for student assessment in grades 6-11, including the cost of lost instructional time, exceeded $1,100 per student. In grades 1-2, testing costs are around $400 per student and in grades 3-5, the annual per-student cost of testing was in the $700 to $800 range.

Conclusion

The Impact of Testing on Instructional Time and the Budget

Given the current fiscal austerity, it is unlikely that taxpayers will finance the longer days or summer school needed to provide the instructional time lost to testing and test prep. The simple alternative is to redirect the time and money devoted to expansive testing to other uses, including but not limited to more instruction. A lot more time for instruction would be available during the existing school day, including restoration of time to subjects that are not tested. Plus, a little money in the budget that is now being spent to purchase and administer tests could be used instead to buy more instructional time or for any purpose, including perhaps spending more money on better tests (such as the Common Core assessments).

Table 5

Impact of Eliminating Tests on Instructional Time and Budget

	Daily Minutes of Added Instruction		Dollars Per Year Available for Other Purposes	
	Midwestern	Eastern	Midwestern	Eastern
K	6	3	$47	$73
1	8	10	$67	$60
2	8	10	$84	$60
3	28	25	$100	$108
4	28	26	$101	$118
5	37	26	$141	$124
6	28	38	$116	$150
7	31	38	$96	$150
8	40	41	$149	$186
9	25	41	$117	$130
10	22	41	$101	$148
11	26	51	$110	$226
12	10	30	$50	$118

*Base on a typical summer school day of 4 hours of instruction.

Table 5 presents the "Daily Minutes of Added Instruction", which comes from the elimination of minutes per day for test-taking, test administration and test prep based on 180 instructional days for Midwestern District and Eastern District. The elimination of test purchasing, licensing and logistics costs is also summed for each grade in each district as "Dollars Per Year Available for Other Purposes".

If testing were abandoned altogether, Midwestern School District could add from 20 to 40 minutes of instruction to each school day for most grades. In grades 6-11, Eastern School District could add almost an entire class period each day. Additionally, in most grades, more than $100 per test-taker would be available for reallocation to purchase interventions, technology, better tests, Common Core staff development, or for any other purpose.

Completely eliminating all testing is both unrealistic and undesirable. However, cutting testing time and costs in half would yield significant gains both to the instructional day and to the budget. It is estimated that the Common Core assessments will cost between $20 to $35 per test-taker more than the current $20 average per test-taker in a typical state (Topol et al., 2012). Cutting the current budget costs for testing and testing logistics in half would easily fund higher-quality Common Core assessments with plenty left over.

These findings suggest the need for more widespread analysis of the instructional costs of testing, particularly at the local district and community level. More importantly, they call for careful consideration of the issues outlined in the recommendations found on pages 7-8. The real costs of testing, along with its purposes, proper uses and instructional impacts—and the voices of frontline educators on these issues—deserve much greater attention in education policy decisions at the federal, state and local levels.

References

Bruno, Robert, Steven Ashby and Frank Manzo (2012). *Beyond the Classroom: An Analysis of a Chicago Public School Teacher's Actual Workday*. School of Labor and Employment Relations, University of Illinois at Urbana-Champaign.

Carlson D., G.D. Borman and M. Robinson (2011). "A Multi-State District-Level Cluster Randomized Trial of the Impact of Data-Driven Reform on Reading and Mathematics Achievement." *Educational Evaluation and Policy Analysis, 33(3),* 378–398.

Chingos, Matthew (2012). *Strength in Numbers: State Spending on K-12 Assessment Systems.* Washington DC: Brookings Institution.

Cornman, Stephen, Jumaane Young and Kenneth C. Herrell (2013). *Revenues and Expenditures for Public Elementary and Secondary Education: School Year 2010-11* (Fiscal Year 2011). National Center for Education Statistics.

Faria, Ann-Marie, Jessica Heppen, Yibing Li, Suzanne Stachel, Wehmah Jones, Katherine Sawyer, Kerri Thomsen, Melissa Kutner, David Miser, Sharon Lewis, Michael Casserly, Candace Simon, Renata Uzzell, Amanda Corcoran, and Moses Palacios (2012). *Charting Success: Data Use and Student Achievement in Urban Schools*. Washington DC: Council of the Great City Schools.

Fryer, Roland (2012). *Learning from the Successes and Failures of Charter Schools*. Washington DC: Hamilton Project, Brookings Institution.

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Bing-ru Teh, Natalya Verbitsky Savitz, Melissa Bowen, Allison Demeritt, Paul Hill and Robin Lake. *Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts*. Cambridge, MA: Mathematica Policy Research, January 2012.

The Gordon Commission on the Future of Assessment in Education (2013). *A Public Policy Statement*. Washington DC: Educational Testing Service.

Hanushek, E. A., and M. E. Raymond. 2005. "Does school accountability lead to improved student performance?" *Journal of Policy Analysis and Management 24*: 297-327.

Henderson, S., A. Petrosino, S. Guckenburg and S. Hamilton (2007). *Measuring how benchmark assessments affect student achievement* (Issues & Answers Report, REL 2007-No. 039). Washington, DC: U.S. Department of Education Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands.

Heppen, Jessica, Wehmah Jones, Ann-Marie Faria, Katherine Sawyer, Sharon Lewis, Amanda Horwitz, Candace Simon, Renata Uzzell and Michael Casserly (2011). *Using Data to Improve Instruction in the Great City Schools: Documenting Current Practice*. American Institutes for Research and Council of the Great City Schools.

Hussar, W., and T. Bailey (2011). *Projections of Education Statistics to 2020* (NCES 2011-026). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Jones, G., B. Jones, B. Hardin, L. Chapman, T. Yarbrough and M. Davis (1999). "The Impacts of High-Stakes Testing on Teachers and Students in North Carolina." *Phi Delta Kappan*, *81*(3).

Lucadamo, Kathleen (2005). "Test Prep Saps Teaching Time." *New York Daily News*.

Marshall , Kim (2006). *Interim Assessments: Key to Successful Implementation*. New York: New Leaders for New Schools.

May, H., and M.A. Robinson (2007) *A randomized evaluation of Ohio's Personalized Assessment Reporting System (PARS)*. Philadelphia: University of Pennsylvania Consortium for Policy Research in Education.

Northwest Evaluation Association (2006). MAP Test Preparation Checklist.

Northwest Evaluation Association (2012). Testing Season Checklist. www.nwea.org/sites/www.nwea.org/files/resources/Testing_Season%20_Checklist.pdf.

Nunnery, J. A., S.M. Ross and A. McDonald (2006). A randomized experimental evaluation of the impact of Accelerated Reader/Reading Renaissance Implementation on reading achievement in grades 3 to 6. *Journal of Education for Students Placed At Risk,* Vol. 1, No. 1.

Pedulla, J., L. Abrams, G. Madaus, M. Russell, M. Ramos and J. Mia (2003). *Perceived Effects of Standardized Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers.* Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Pennsylvania Department of Education (2011). *2011-2012 PSSA/PSSA-M Assessment Handbook*. Pennsylvania Department of Education, Bureau of Assessment and Accountability, Division of Assessment.

Quint, J., S. Sepanik and J. Smith and MDRC (2008) *Using student data to improve teaching and learning: findings from an evaluation of the Formative Assessments of Students Thinking in Reading (FAST-R) Program in Boston elementary schools.* Retrieved from ERIC database (ED503919).

Ross, S. M., J. Nunnery and E. Goldfeder. (2004). *A Randomized Experiment on the Effects of Accelerated Reader/Reading Renaissance in an Urban School District.* Memphis, TN: Center for Research in Educational Policy, University of Memphis.

Roza, Marguerite and Karen Hawley Miles (2008). *Taking Stock of the Fiscal Costs of Expanded Learning Time.* Washington, DC: Center for American Progress.

Silva, E. (2007). *On the Clock: Rethinking the Way Schools Use Time.* Washington, DC: Education Sector.

Slavin, Robert, Alan Cheung, GwenCarol Holmes, Nancy Madden and Anne Chamberlain (2013). "Effects of a Data-Driven District Reform Model on State Assessment Outcomes," *American Educational Research Journal*, Vol. 50, No. 2.

Tepper Jacob, Robin and Susan Stone (2005). "Teachers and Students Speak," *Education Next,* Vol. 5, No. 1.

Texas Federation of Teachers (2005). Texas Federation of Teachers Survey on the Impacts of TAKS.

Topol, Barry, John Olson, Ed Roeber and Pat Hennon (2012). *Getting to Higher-Quality Assessments: Evaluating Costs, Benefits, and Investment Strategies*. Assessment Solutions Group.

Yeh, S. (2008). The Cost-Effectiveness of Comprehensive School Reform and Rapid Assessment. *Education Policy Analysis Archives* (13), http://epaa.asu.edu/epaa/v16n13/.

——— (2010). The Cost-Effectiveness of 22 Approaches for Raising Student Achievement. *Journal of Education Finance*, Vol. 36, No. 1.

Ysseldyke, J. and D.M. Bolt, D. M. (2007). Effect of Technology-Enhanced Continuous Progress Monitoring on Math Achievement. *School Psychology Review, 36*(3), 453–467.

Ysseldyke, J. and S. Tardrew. (2007). Use of a Progress-Monitoring System to Enable Teachers toDifferentiate Math Instruction. *Journal of Applied School Psychology, 24*(1), 1–28.

Zellner, Michael, Anthony Frontier and Denise Pheifer (2006). "What Are NCLB's Instructional Costs?" *Educational Leadership*, Vol. 44, No. 3.

APPENDIX
Research on Time Used for Test Preparation

Although there is rich literature on test preparation, very little research exists on the *time* students spend preparing for tests. Our study creates a rubric (Table 4) for estimating test prep time based on findings from two high-quality studies.

Center for the Study of Testing, Evaluation, and Educational Policy. The 80-item survey of teachers (Pedulla et al., 2003) sought to ascertain teacher' attitudes and opinions about state-mandated testing programs and then analyzed the data according to the nature of the consequences or stakes attached to their state test results. For districts, schools or teachers, *high stakes* refers to state-regulated or legislated sanctions of significant consequence such as accreditation, financial rewards or placing a school in receivership. The *low-stakes* category included states with testing programs that did not have any known consequences attached to test scores. If the stakes attached to the state tests did not meet the criteria of either the high- or low-stakes definitions, states were placed in the moderate category.

Responding to the question, "Approximately how many class hours PER YEAR do you spend preparing students specifically for the state-mandated test (i.e., teaching test-taking skills)?" teachers responded as follows:

> In high-stakes states, 13 percent of teachers spent 21-30 hours on test prep, and 44 percent devoted more than 30 hours to test prep.
>
> In low-stakes states, 7 percent of teachers spent 21-30 hours on test prep and 44 percent devoted more than 10 hours to test prep.

The study found that test prep practices vary with the stakes attached to testing. Chosen from a list of test preparation practices, the largest differences between teachers in high-stakes and low-stakes states are:

> "I provide students with items similar to those on the test" (75 percent compared to 54 percent);
>
> "I provide test-specific preparation materials developed commercially or by the state" (63 percent compared to 19 percent); and
>
> "I provide students with released items from the state-mandated test" (44 percent compared to 19 percent).

Teachers in high-stakes states were much more likely to report that their districts required or recommended summer school for low-scoring students (43 percent compared to 8 percent), or retain students in grade (25 percent compared to 3 percent).

Consortium on Chicago School Research: The centerpiece of Chicago's high-stakes testing program implemented in the late 1990s was a set of minimum test score standards on the reading and mathematics sections of the Iowa Test of Basic Skills for students in grades 3, 6 and 8. Students who did not meet the test score cutoffs at the end of the school year were required to participate in a special summer program.

Survey data collected by the consortium (Tepper Jacob and Stone, 2005) showed that teachers defined test preparation narrowly. Teachers *did count* giving practice tests and teaching test-taking strategies as test prep. Teachers *did not count* activities aligning content to the test as test prep. Half of the teachers, for instance, said they had aligned the content of their curriculum with the content of the test, but only 4 percent included curriculum content in what they considered "test preparation." Based on interviews, teachers viewed test preparation as the most effective way to improve test scores. Teachers used a variety of approaches to test prep emphasizing, above all, test simulation and familiarization with the layout and format.

Data collected from teacher surveys before implementation of the high-stakes social promotion policy in 1994, and after five years of implementation in 1999, yielded teacher estimations of time they spent on test preparation:

Test prep time doubled with the high-stakes pressure. Before high-stakes testing, teachers spent 10.5 hours on test preparation compared with 21 hours five years into high-stakes testing. Teachers in the lowest-performing schools increased test preparation from 14 hours prior to the high-stakes testing policy to 32 hours five years into implementation.

Impact was greater in high-stakes grades. Before the high-stakes testing regimen, 30 percent of teachers in high-stakes grades (grades 3, 6 and 8) spent more than 20 hours on test prep. Five years into implementation, 65 percent of teachers spent more than 20 hours. Among third-grade teachers, test prep moved from an average of 14 hours to 31 hours.

Appendix Table A

Direct Cost of Assessments and Licensing, Midwestern School District

	K	1	2	3	4	5	6	7	8	9	10	11	12
Locally Adopted Assessments													
NNAT2 – Naglieri Nonverbal													
DRA	$14		$14				$14		$14				
Stanford Achievement Test	$22	$22	$22	$22	$22	$22	$22						
21st Century Skills Test		$29	$29										
ACT EXPLORE, or PLAN						$6			$6				
Career/technical EOC								$9	$9	$11	$11	$34	$22
Total Local	$36	$51	$65	$22	$22	$28	$36	$9	$29	$11	$11	$34	$22
Interim/Benchmark Assessments													
ACUITY – Math, Comm. Arts				$20	$20	$20	$20	$20	$20				
ACUITY – Science						$10			$10				
Scanton – Math, Comm. Arts								$14	$14	$14	$14	$14	
Mock End of Course										$14	$14	$14	$14
Total Interim/Benchmark				$20	$20	$30	$20	$34	$44	$29	$29	$29	$14
State-Mandated Assessments													
Math, Comm. Arts				$25	$25	$25	$25	$25	$25	$25	$25	$25	
Science						$10			$10	$10			
Social Studies										$10	$10		
Total State-Mandated				$25	$25	$35	$25	$25	$35	$45	$35	$25	
Total Local, Interim and State	$36	$51	$65	$67	$67	$93	$81	$68	$108	$85	$75	$88	$36

Appendix Table B

Direct Cost of Assessments and Licensing, Eastern School District

	K	1	2	3	4	5	6	7	8	9	10	11	12
Locally Adopted Assessments													
GRADE (Pearson)				$14	$11	$11	$11	$11	$11	$11	$11	$11	
Kindergarten Assessment	$15												
DIBELS	$25	$25	$25	$25	$25	$25							
SRI Reading							$15	$15	$15	$15	$15	$15	$15
TerraNova	$25	$25	$25										
CBA Math		a	a				a	a	a	a	a	a	a
CBA Reading							a	a	a	a	a	a	a
CBA Science							a	a	a	a	a	a	a
CBA Social Studies							a	a	a	a	a	a	a
PSAT											$14	$14	
Career/Technical EOC													$22
Total Local	$65	$50	$50	$39	$36	$36	$26	$26	$26	$26	$40	$40	$37
Interim/Benchmark Assessments													
CBA Math				b	b	b							
CBA Reading				b	b	b							
CDT Math, Reading, Science							$20	$20	$20	$20	$20	$20	
State-Mandated Assessments													
End of Course										$10	$10	$10	$10
Reading, Math				$25	$25	$25	$25	$25	$25			$25	$25
Science					$10				$10			$10	
Writing						$10			$10			$10	
Total State-Mandated				$25	$35	$35	$25	$25	$45	$10	$10	$55	$35
Total Local, Interim and State	$65	$50	$50	$64	$71	$71	$71	$71	$91	$56	$70	$115	$72

Appendix Table C

Estimated Hours of Test Preparation Time, Midwestern School District

	K	1	2	3	4	5	6	7	8	9	10	11	12
Locally Adopted Assessments													
NNAT2 – Naglieri Nonverbal	0		0				0		0				
DRA	0	0	0	0	0	0	0						
Stanford Achievement Test	10	12	12										
21st Century Skills Test						7			7				
ACT EXPLORE, or PLAN								0	0	0	0	0	
Career/Technical EOC													10
Total Local	10	12	12	0	0	7	0	0	7	0	0	0	10
Interim/Benchmark Assessments													
ACUITY - Math, Comm. Arts				32	32	32	32	32	32				
ACUITY- Science						5			5				
Scanton - Math, Comm. Arts								12	12	12	12	12	
Mock End of Course										12	12	12	12
Total Interim/Benchmark				32	32	37	32	44	49	24	24	24	12
State-Mandated Assessments													
Math, Communiction Arts				30	30	30	30	30	30	20	20	20	
Science						5			5	5			
Social Studies										5	5		
Total State-Mandated				30	30	35	30	30	35	30	25	20	
Total Local, Interim and State	10	12	12	62	62	79	62	74	91	54	49	44	22

Appendix Table D

Estimated Hours of Test Preparation Time, Eastern School District

	K	1	2	3	4	5	6	7	8	9	10	11	12
Locally Adopted Assessments													
GRADE (Pearson)				12	12	12	12	12	12	12	12	12	
Kindergarten Assessment	0												
DIBELS	0	0	0	0	0	0							
SRI Reading							12	12	12	12	12	12	12
TerraNova	10	10	10										
CBA Math		12	12				7	7	7	7	7	7	7
CBA Reading							7	7	7	7	7	7	7
CBA Science							7	7	7	7	7	7	7
CBA Social Studies							7	7	7	7	7	7	7
PSAT											0	0	
Career/technical EOC													10
Total Loccal	10	22	22	12	12	12	52	52	52	52	52	52	50
Interim/Benchmark Assessments													
CBA Math, Reading				32	32	32							
CDT Math and Reading							32	32	32	32	32	32	
Total Interim/Benchmark				32	32	32	32	32	32	32	32	32	
State-Mandated Assessments													
Ennd of Course										40	40	40	40
Math, Reading				30	30	30	30	30	30			20	
Science					5				5			5	
Writing						5			5			5	
Total State-Mandated				30	35	35	30	30	40	40	40	70	40
Total Local, Interim and State	10	22	22	74	67	79	114	114	124	124	124	154	90

A Union of Professionals

American Federation of Teachers, AFL-CIO
555 New Jersey Ave. N.W.
Washington, DC 20001
202-879-4400
www.aft.org