

# Chapter 7

## Assessment Measures

### Highlights of Findings

**The SAT-9 is pervasive, but its influence is viewed as highly problematic by many districts and schools.**

Not surprisingly, almost all districts cited this test as one of their accountability measures. As a mandatory, high-stakes test, the SAT-9 exerts considerable influence at the school level — affecting what professional development opportunities are offered and how classroom time is spent on test preparation. However, the SAT-9 is viewed by district- and school-level personnel as inherently flawed for the following reasons:

- The SAT-9, in its current form, is not aligned with state standards.
- The emphasis on the SAT-9 promotes “teaching to the test” — that is limiting the curriculum to what is on the test.
- The “augmented” standards-based items included in the SAT-9 are considered overly difficult and unfair.
- The requirement that all students be tested in English is unfair.
- Testing takes valuable time from instruction.

With regard to the Public School Accountability Act of 1999, local and school personnel expressed grave concern with the Academic Performance Index’s over-reliance on SAT-9 results, particularly as the test is not fully aligned to state standards. Some believe that the emphasis on the SAT-9 may drive the education system in harmful ways.

**Most districts use and value multiple measures, but fear they will be abandoned because of new state accountability requirements.**

Consistent with state requirements that existed prior to the passage of the state’s new accountability law, most districts report using a wide array of multiple measures, including:

- writing samples
- grades
- teacher judgments, and
- district criterion-referenced tests.

Many district personnel express concerns that the state’s increasing emphasis on the SAT-9 as the sole indicator of student performance will lead to the demise of multiple measures.

As part of the statewide Standards-Based Accountability System, the state required school districts submitting Consolidated Applications for Categorical Funding to have at least one achievement measure each in language arts and math for every grade level. In order to comply with federal mandates for Title I funding, districts also had to use multiple measures of assessment for at least one grade level in each of three specified grade spans. In addition, districts were required to establish student performance standards showing how the multiple measures would be combined to determine student proficiency. These reporting requirements first went into effect in 1997. In 1998, the state established statewide criteria for combining multiple measures and required that the Standardized Testing and Reporting (STAR) test — the Stanford Achievement Test, 9<sup>th</sup> Edition, Form T (SAT-9) — be one of the measures used (McKenna, 1998; Fausset, 1998).

For the purposes of this study, the overarching research question on the topic of assessment was:

***What is the nature of local assessments for standards-based accountability and how are they used to determine whether a student has met or has not met the local standards?***

The Stanford-9

*The SAT-9 played a large part [in influencing mathematics instruction at this school] — fortunately and unfortunately. You want to teach the kids what they will be tested on.*

—Principal

**The SAT-9 is the most commonly used assessment measure. It has made a significant impact on schools and teachers and in some places appears to drive curriculum and instruction.**

Given that the use of the SAT-9 is required by the state, it would be surprising to find many districts that did not include it as one of their assessment measures. Indeed, nearly all survey respondents reported that the SAT-9 is used at one or more grade levels for accountability purposes. Among district accountability plans, 90.5 percent of the 190 CDE-rated plans, and 91.7 percent of the 36 independently reviewed plans, also cited use of the SAT-9 as an accountability measure. Moreover, among the CDE-rated plans, only 6 percent were found not to include results from the STAR test for all students in grades 2–11, and only 0.5 percent (i.e., 1 district out of 190) apparently did not use the SAT-9 at all for accountability reporting.

**A few district survey respondents commented about the effects of the SAT-9 on instruction:**

*Because of accountability, school staffs are modifying their instruction to address weaknesses on SAT9.*

*SAT9 has created a high apprehension level on teachers and principals and community; some positive classroom results are observable.*

*Use of NRTs in a public way has increased the high stakes nature of assessment and resulted in focused attention that is new.*

School-level personnel confirmed that they are aware of the SAT-9 and its importance as an accountability measure. Numerous principals and teachers spoke of such things as “living and dying by the test scores,” focusing professional development efforts on improving test scores, pacing instruction so that teachers can “strategically prepare” the students for standardized tests, and “anxiously awaiting” the SAT-9 results. (Interviews were conducted before the scores were released.) One principal explained that “the SAT-9 has been the catalyst” for changes occurring in her school; “Other state policies,” she continued, “have had nowhere near the same level of influence.”

Indeed, in answer to the question, “Did you do anything special to help your students prepare for this year’s SAT-9?,” the vast majority of teachers interviewed answered in the affirmative. A few of the teachers focused on basic skills or on particular content areas as part of this preparation. More common responses, however, included work on “test-taking skills” (for example, in taking multiple-choice tests) and the administration of practice tests. Several teachers mentioned the use of test-preparation booklets/materials, although in more than one instance, these materials had not arrived in time to be used this year.

The amount of time spent specifically on SAT-9 preparation was variable. Roughly one-third of the teachers who were interviewed said they’d spent two to three weeks; about another third said one to two months or one day per week all year long. A few teachers did say that they had worked on SAT-9 preparation all year.

At many of the schools visited, the influence of the SAT-9 goes beyond preparing students to take the test and extends into the realm of shaping the curriculum itself. Without prompting, many teachers mentioned the SAT-9 in their answers to questions about their “general approach” to teaching mathematics or about documents and policies that they felt had had an impact on their teaching. “The thing that jumps to mind is the STAR-9 testing,” replied one teacher; “the greatest impact comes from the Stanford-9 and [another assessment used in the district],” stated another. Responses such as these, along with “preparing students to take standardized tests,” were fairly typical.

Moreover, several principals stated unequivocally that the SAT-9 will “drive the way we teach” or had already done so. Some teachers, as well, made comments about the influence of the SAT-9 over their curriculum or their instructional practices. “The test influences what I teach,” explained one teacher, “I try to cover all the areas that will be on the test,” she continued. In a different district, another teacher remarked that after the students had taken this year’s SAT-9, she asked them what they did not know on the test; they indicated geometry, so next year she intends to bring that in earlier. More generally, this same teacher stated, “If the SAT-9 is a test of skills, not theory, then we might as well continue to teach that way.”

**District- and school-level personnel expressed grave concerns about overreliance on the SAT-9, especially given a perceived lack of alignment between the test, the standards, and the curriculum.**

*The normed SAT-9 test has had a negative effect on teachers, students, and parents because the test is not aligned to daily instruction. Therefore teachers and students are defeated before they begin.*

—District Survey Response

*The new adoption for the district — there’s an obvious philosophy behind it that it should be hands-on...My biggest complaint with the hands-on is that [students are] not tested that way. It’s like they [the district] want us to use hands-on materials, but then they test us in a much more traditional way...Regarding the district and the state, teachers are getting mixed messages about hands-on versus seatwork. I don’t get a consistent message. No one fully explains to you how you’re supposed to prepare kids for tests.*

—Teacher

Many districts’ questionnaires contained remarks expressing serious concerns about the SAT-9 and the accountability emphasis placed on it. Terms such as “overreliance” and “overemphasis” turned up again and again in respondents’ comments about the SAT-9. The following comments, given in response to various open-ended questions, were typical:

*The overemphasis on a norm referenced test statewide is very problematic.*

*Overreliance on norm-referenced test data [has hindered our district].*

In particular, many districts take issue with the centrality of a norm-referenced test in a standards-based system. The use of an NRT as part of a standards-based approach, wrote one respondent, “is somewhat unreasonable.” Several survey respondents commented that

**the SAT-9's lack of alignment to the state standards and to curriculum and instruction is deeply problematic, given the importance of the SAT-9 for accountability:**

*Standardized tests not aligned to standards cause problems in the implementation of standards.*

*Over reliance on SAT9 (NRT) (one measure that is norm referenced instead of criterion referenced and isn't matched to standards).*

*The district is concerned about...the use of a nationally normed test to rate performance on state standards.*

*Student achievement is being gauged by a test NOT in alignment with State Standards.*

Some district staff made similar comments in interviews. For example, one Director of Research and Evaluation acknowledged that there had been a major change in district classrooms of late, but he attributed this change to the STAR test *rather than* to the content standards. An official in a different district expressed concerns about “mixed messages” from the state related to the use of the content standards and the SAT-9. He said that he finds it particularly difficult to explain to parents how a standards-based system works in the context of using the SAT-9 to assess achievement. He also mentioned a concern that reliance on the SAT-9 will divert instruction from the broader, more rigorous content standards to the more narrow basic skills measured by the test. Similarly, one district survey respondent wrote, “C & I practices are harmed by forcing a narrowing of learning outcomes to testable areas.”

At the school level, as well, many people commented on a perceived lack of alignment between the test and the content standards. For example, one principal said that “we have no measure” for determining if a student meets the district standards, implying that the SAT-9 does not serve this purpose. Another principal mentioned that there had been “some resistance to the SAT-9 because it is not aligned with the standards.” As an example, she pointed out that in social studies, seventh graders are tested on U.S. history, but they do not take U.S. history until eighth grade. She expounded further with regard to mathematics:

*Do [the district math] standards align with the standardized tests that [students] have to take?  
No. They don't. And that's very frustrating for math teachers. What we're teaching and when we're teaching it, and when they take the standardized tests and they see that something is on there that they haven't taught yet — it's very frustrating. ... Aligning our state testing with our standards is really important, so we don't have that frustration.*

Another principal in the same district said that there had been much anxiety in her school over the STAR program; she said that the teachers were worried that the kids were being

tested on topics not taught. A principal in a different district made a similar comment, about teachers seeing “a discrepancy” between things on the test and things that are taught.

Several teachers, too, remarked on the lack of alignment between the test, the curriculum, and the standards. “The Stanford-9 test material is not in our curriculum!” bemoaned one teacher. Another spoke of how the SAT-9 was a “more traditional” approach that does not mesh with the curriculum. Other teachers mentioned that “it would be helpful” to have assessments that were aligned to the new standards. “The test doesn’t assess what’s going on here,” stated one teacher; “The SAT-9 is not a good judge,” said another.

**At the classroom level, many teachers feel that they are being compelled to “teach to the test,” a particular problem if the test is not aligned with the standards.**

As suggested by the remarks from people who say that the test is driving curriculum, it appears that many schools and teachers are adapting instruction to fit the test. But many teachers strongly object to the idea of “teaching to the test,” though they fear that ultimately they will be asked to do so. As one teacher put it:

*I don’t really believe in teaching to a standardized test. I think that too much importance is placed on them. At the same time, I’m tugged in that direction, because everybody thinks it’s important....So, I have to honor it.*

Another teacher stated strongly, “The SAT-9 is going to have a negative impact. It really controls teaching and what is taught.” Other teachers particularly objected to the idea of “teaching to a test” that is not aligned with the standards:

*I get the impression from the state government that we need to teach to the test. I mean, who cares about content anymore in the math class? We teach to the test. Because now they [the state government] are offering extra money tied to teachers whose test scores are high. And, so that speaks very loudly that...it doesn’t matter about the content, let’s teach to the test....I’m not going to, but that’s what I’m hearing, and I’ll bet you that, in time, the department will force me to do that....I think the standardized test that we have to take gets in the way. Because it forces me to teach to the test, instead of teaching to what the standards are.*

*There’s a lot of pressure to make sure students perform well on [the SAT9]. And personally, I think if the curriculum is strong and you teach the curriculum, then you don’t have to worry about the individual test. But, I’m kind of shouting out in a field by myself on that. Or, at least, there are a lot of teachers shouting out there, and other people aren’t listening. And I just fear that we’re moving too much toward teaching to a test. It’s not ever been stated that way, but I think it’s moving in that direction. I avoid it [teaching to the test], thinking that the strength of the*

*curriculum will do the job. And, I don't know what I'll have to do if the results aren't good, and I have to revise what I do. Because, I think, then the task is, change the curriculum...I think the problem we have right now is that the test and the curriculum are based on different standards, and they haven't brought them in line. And I'd like to see the test follow the curriculum — or, decide what the curriculum should be, establish the statewide standards, or national standards, or whatever the heck we're going to use, and then make sure the test follows that. And not the other way around. I don't want a curriculum chasing the test. I want the test to match the standards. And I don't think we're anywhere near there yet....*

Some principals also expressed concerns about curriculum driven by assessment. One principal commented that looking at test scores might help improve the scores, but that this did not necessarily mean improving the curriculum. Another principal said that she worries that as teachers teach more and more narrowly to the tests, important things are getting left out of children's education. For example, until the state starts testing science and social studies, she speculated, schools will concentrate almost exclusively on math and language arts. Previously, this principal remarked, she would have felt accountable to parents to give children a well-rounded education, but she now feels accountable to the district (who in turn is accountable to the state) to provide high scores. She thought that this sometimes gets in the way of giving the kids the best possible education.

Interestingly, some of the same concerns were expressed during interviews by officials within the California Department of Education. Department staff observed that the STAR test has the very real potential to cause schools to narrow curriculum and instruction in order to maximize student performance on the test. Some argued that raising test scores on this one standardized test is not synonymous with improving education, as it is possible to raise test scores without improving student learning. The comment was also made that no matter how prominently state standards are endorsed, unless they are assessed, it is unlikely that they will be taken seriously in many school districts.

**Although the augmented portion of the SAT-9 may be more aligned with the standards than the rest of the SAT-9, the use of the augmented test in spring 1999 caused considerable anguish among districts, schools, teachers, and students.**

*This year the state augmented test — there were things on there that people thought were really not being taught. So does that mean they change their curriculum? I don't think so...there's a lot of resistance to that.*

—Principal

At least in theory, the use of the new “augmented,” standards-based sections of the SAT-9 may alleviate some of the concerns that people have about lack of alignment between the test and the standards. However, it appears that considerable progress remains to be made with the use of these new sections. According to sources within the California Department of Education, some of the augmented items failed to meet technical standards of validity and reliability. And at least one district survey respondent complained that “poor augmented test items” had been a hindrance in the development of an accountability system.

Several school-level staff and teachers objected to the augmented sections of this year’s test as well. One principal remarked, “the augmentation portion was a bust”; she said that the test “set the students and teachers up because the expectations were not matched by what students found on test.” Another principal said merely that the augmented test had “caused confusion.” And the following remarks were made by teachers in two different districts, the first one an elementary school teacher and the second one a middle school teacher:

*I was really upset by the augmentation test. The students were asked to work with negative integers. I didn’t teach them that.*

*The SAT9 tests a lot of stuff that they haven’t even learned...The problem is that we’re supposed to be aligned with the state test. And so, that means basically we need to advance all our students before they’re ready....The 7<sup>th</sup> graders had to take this test, the STAR test...While they were taking it, I could just see the frustration on their faces, and I was like, what’s going on? So I grabbed a copy of the test. I started looking at it; I was like, oh my gosh, they’re so frustrated because this is the stuff I’m teaching my 8<sup>th</sup> graders right now, but my 7<sup>th</sup> graders haven’t even seen this material yet.*

This teacher said that as a result of this experience, next year he plans to move content down from the 8<sup>th</sup> grade to the 7<sup>th</sup> grade to the “best of his ability.” Similarly, another teacher said, “I don’t believe in teaching to the test but it’s not fair for a child not to have exposure to what’s on the test.” She indicated that next year, she will add new topics to her curriculum — those on the augmentation test — so that students have exposure to them.

In this way, then, the use of the augmented portions of the SAT-9 may indeed be having the effect desired by the state: they are spurring teachers to teach particular content at levels they otherwise would not have. To the extent that this content is indeed aligned with the standards, then the test is encouraging standards-based instruction. As one principal put it, “[The augmented test] has really been an issue with our math teachers, because they feel that it’s out of reach of most students. But maybe that’s the purpose of it: make it within reach.”

The time at which the test is administered, however, also plays an important role on how much of the content students have covered.<sup>1</sup> One teacher stated that although the test was given in the early spring, it focused on the last third of the year’s curriculum, and the class simply “hadn’t gotten to a lot of those topics yet.” Another teacher, interviewed toward the end of the school year, said that her class had covered several more standards since the test was given, as a result of the way the book was set up. She hypothesized that if her students could “take the [augmented] test today, they could get at least ten more right.” As it was, however, she stated, “The SAT9 was extremely frustrating — it was a whole week of upset and tears” for her students, whom she said are among the best at her school. She teaches five gifted classes.

**Many district and school staff object to the requirement that *all* students, especially English language learners, take the SAT-9.**

About 40 percent of the 190 plans reviewed by CDE used as one of their multiple measures a primary language NRT. Nevertheless, *all* students are required by the state to take the SAT-9 in English — including English language learners, even if they are also taking a primary language NRT.

Several districts raised concerns about this requirement. In a general open-ended question on the survey about what state-level practices had hindered districts in developing and implementing their accountability systems, the following responses were among those given:

*English SAT-9 for ELs*

*Testing non-English speaking LEP students with SAT-9 who are not Spanish speaking.*

*Requiring non-English speakers to take a test in English*

*Requiring all LEP to take SAT-9*

*The requirement for non-English speakers to take the test in English.*

One survey respondent elaborated a bit on how the policy had been problematic for her district:

---

<sup>1</sup> This may be a particular issue for districts or schools that have year-round calendars. The Director of Research and Evaluation in one large district said that his district’s track system does not fit well with the state’s mandated testing window, leaving students in some tracks with up to 40 days less instruction before being tested.

*Requirement to test all Limited English Proficient students in the SAT-9 yield invalid scores (student & school) and thus makes it difficult for us to compare schools since the student populations tested are so different.*

**This survey respondent was not the only person to express concern that the SAT-9 does not take into account differences between schools or students. A similar remark was made by a principal at an overcrowded urban school in which 58 out of the 79 teachers on staff last year were “newly installed”:**

*The measures of accountability (Stanford-9) do not account for the context of the school. There isn't much parent support because most parents have little or no education. The average teacher expectancy is 3 years. There is a high teacher turnover. The accountability to the district through the Stanford-9 doesn't measure progress within a particular context. The playing field is not level.*

**A few teachers, as well, voiced concerns that the SAT-9 was not sensitive to matters of equity. One stated, “The state constraints are based on tests only with no focus on reality. The SAT-9 was designed for upper middle class with no ethnic diversity.” Another teacher was more concerned about the repercussions for *her* of certain students taking the SAT-9:**

*Some kids [e.g., English language learners, special education students] shouldn't have to take the standardized test, and if they still have to, and those scores are counted into my scores, into my teaching, and I'm held accountable for that, then I kind of have a problem with that....And the other factor is transience. I mean, there's a lot of kids who bounce from school to school to school, and if I have not taught them all year long, it doesn't seem fair to me to be held accountable for them.*

## Grades, Teacher Judgment, and Writing Samples

**Other frequently used assessment measures include report card grades, teacher evaluation and judgement, and writing samples.**

On the district survey, more than 80 percent of respondents said that their districts use report card grades at one or more grade levels for accountability purposes, and 55 percent use some form of teacher evaluation/judgment. The CDE review of 190 plans found that 65.3 percent of districts were using grades as a measure; the independent review of 36 plans tallied 80.6 percent using grades and 13.9 percent using teacher evaluation/judgment. Also frequently mentioned were writing samples (86.3% of survey districts, 46.8% of CDE plans, 44.4% of independently reviewed plans).

Such widespread use of these types of measures may be cause for concern, given that they seldom are tested for reliability and validity across classrooms and schools. Indeed, nearly one-quarter of districts reported on the survey that they do not use *any* process to ensure that class grades are consistent and comparable across district schools. Among districts that do have such processes, 48.1 percent said that the district issues a grading policy or guidelines. The rigor and implementation of such policies, however, may be suspect; in some districts, these “policies” may be little more than pieces of paper. But in one interview district, where teacher judgment is one of three measures used to assess student mastery of standards, there have been significant efforts to obtain consistency in the use of this measure. Teachers are provided with training on how to judge student work within a framework of four levels of student proficiency: advanced, proficient, approaching proficiency, and below proficiency.

Another process cited by many survey respondents (47.3%) for ensuring consistency and comparability of grades was the comparison of student achievement on multiple assessment measures for triangulation of data. However, this, too, may be problematic, or even undesirable. One Director of Research and Evaluation had the view that multiple measures should be measuring different things; otherwise there is no reason to use more than one measure. He supports the use of grades as one of the multiple measures and cited evidence that teachers base grades on the kinds of things he feels are important in assessment. In his district, the correlation between grades and test scores has not been very high (around 0.3 or 0.4); this is viewed as evidence that grades and tests are measuring different things. He acknowledged that a high stakes accountability system might lead to grade inflation, but said that he has not yet seen this happen.

One possibly promising direction in terms of the use of grades as a measure in a standards-based accountability system lies with the notion of standards-based report cards. On the survey, a few districts wrote that they are in the process of developing a standards-based report card, and a couple of the interview districts mentioned this as well. In fact, the district discussed just above — the one that values grades as a measure — is working on aligning student report cards with standards.

Another interview district was farther along in the use of standards-based report cards. New grade books were developed to capture student progress in meeting standards at the elementary school level, and elementary schools are currently using standards-based report cards. These report cards have been called “pilot versions” for the past four years in order to avoid too much controversy, but, according to parent surveys within the district, 93 percent of parents are in support of the standards-based report cards. According to the district staff and principals, teachers were wary of the standards-based approach at first but for the most part have since come to embrace it. The district plans to tailor the report cards for quarterly reporting, after determining where students should be in meeting standards at various points during the year.

## Criterion-Referenced Assessments

**Many districts have developed or purchased their own criterion-referenced assessments — a difficult, expensive, and time-consuming process.**

Many districts also reported using criterion-referenced assessments both in English/language arts (55.7% of survey districts, 55.8% of CDE plans, 52.8% of independently reviewed plans) and in mathematics (57.3% of survey districts, 54.7% of CDE plans, 50% of independently reviewed plans).

Three of the districts in which researchers conducted interviews, including the two smaller ones, developed or purchased assessments to use as one of their multiple measures. For example, one of these districts developed an assessment system for reading based on an existing model; it is an in-depth diagnostic tool given three times a year to all students in grades 1–8, and it covers eight reading subcategories. For mathematics, this district bought for diagnostics a commercial assessment that is aligned to the SAT-9. The results of the diagnostics are supposed to be used by teachers to help guide instruction for individual students and are also used in parent-teacher conferences.

Several survey respondents seemed proud of locally developed assessments and indicated that the use of these assessments is having a positive impact on classroom instruction:

*District-made K–3 reading assessments are guiding instructional practices in the classroom.*

*District Writing Assessment is scored with a rubric that is now included in classroom instruction. Teachers report positive impact.*

*What is tested is taught. We've taken great care to develop assessments. Staffs work together to analyze results of assessments. We can now hold teachers accountable.*

*Assessments have been developed or identified and purchased that teachers are using to monitor student achievement in reading and writing K–5, and they are using assessment results to plan instruction.*

However, district interviews revealed that the development of local assessments can be a difficult, expensive, and time-consuming process. Staff from one district, for instance, mentioned that the district has spent a lot of time developing multiple measures and has had great difficulty in coming up with valid, reliable tests. The district eventually bought a test bank from a publisher, but it still needs to develop assessments for each grade. (They would like to be able to hire a full-time Assessment Director to develop the measures,

oversee the student data tracking system, and evaluate the data.) Staff from this district also mentioned that committees of district staff, principals, and teachers have been scrambling to develop a mathematics assessment for each primary school grade. They did not have the capacity to create what they felt was a good, reliable and valid test, but they felt they had no choice because it was required for Consolidated Application reporting.

Conversations with representatives from rural districts indicate that such districts are particularly concerned about their capacity to develop assessment and data analysis systems. Some districts feel it is beyond their capacity to develop assessment systems, and would like to see the state offer an array of approved, valid, and reliable options.

Although locally developed assessments may have some of the same types of reliability and validity problems that grades and teacher judgment have, most districts are aware of these problems. Indeed, 111 of 114 survey districts (97.4%) reported that they provide professional development for teachers on the use and/or scoring of locally developed assessment measures. In one district that uses local assessments, the Director of Research and Evaluation reported in an interview that assessments are scored both by the classroom teacher and centrally by the district; any discrepancies are given to a third reader.

An official from a different district noted that designing a reliable and valid test was not the only problem. In his opinion, getting consensus on what should be measured was even more difficult. Although this issue sparked a productive community-wide debate on what students should know and how that should be measured, this debate prolonged an already-lengthy development process. Because the district had to meet state reporting deadlines, the multiple measures were implemented before they were completed to everyone's satisfaction. Though the measures continue to be refined, the district now has to overcome the negative reaction of the teachers, parents, and students, who rebelled against the hastily designed assessments. In addition, since the assessments change each year, the district has not been able to establish baseline data against which to benchmark progress.

### **Testing takes time away from instruction.**

Another area of concern with regard to assessment is the amount of class time needed to administer and prepare for tests. A comment on one district's questionnaire was that "Teachers are concerned about the amount of classroom time being lost to assessment," and this was generally supported in interviews with school-level personnel. One principal, for example, was concerned that too much time was being devoted to testing, and that it was cutting into instruction time. And a teacher in a different district said he thought that the district assessment, given three times over the course of the year, "was a little much":

*We lost three instructional days, plus whatever preparation we were doing for it. And then also, it took some time to grade the papers, all that kind of stuff, which took away from my preparation*

*time as well....So I thought it was a little much...to do three of them; I felt it would be better if it was just one.*

Some teachers said that they had stopped what they were doing in order to prepare students for assessments (including the SAT-9), and a few of them resented having to do this. As one teacher put it, “It [test preparation] slowed me down with respect to my regular instruction.”

Also, some teachers voiced concerns not only about the amount of time required to prepare for and to give assessments, but the particular scheduling of these assessments, as indicated by the following comments from two teachers at one school:

*I mean, it's really hard, because, like, we'll get a test coming up, a [district] performance-based assessment test, coming up, and I'll look at it, and I'll go, "Oh, gee, we haven't even covered this yet." So I'll have to stop what I'm doing, cover this material, so that they can do well on the performance-based assessment test. And then go back to my regular material.*

*The district has had...performance-based assessments that we had three times this year...And I have no trouble doing performance-based assessments, but when it comes from the district, it doesn't necessarily fit with what you're doing at the time. I'd rather have an assessment that goes along with what they [students] are doing...It was like, just take this chunk out of time, and do this thing that's not associated with what you're teaching.*

**Districts report that they are working to align local assessments with their district standards. About 50 percent of surveyed districts report significant alignment of local assessments and district standards in English/language arts and mathematics.**

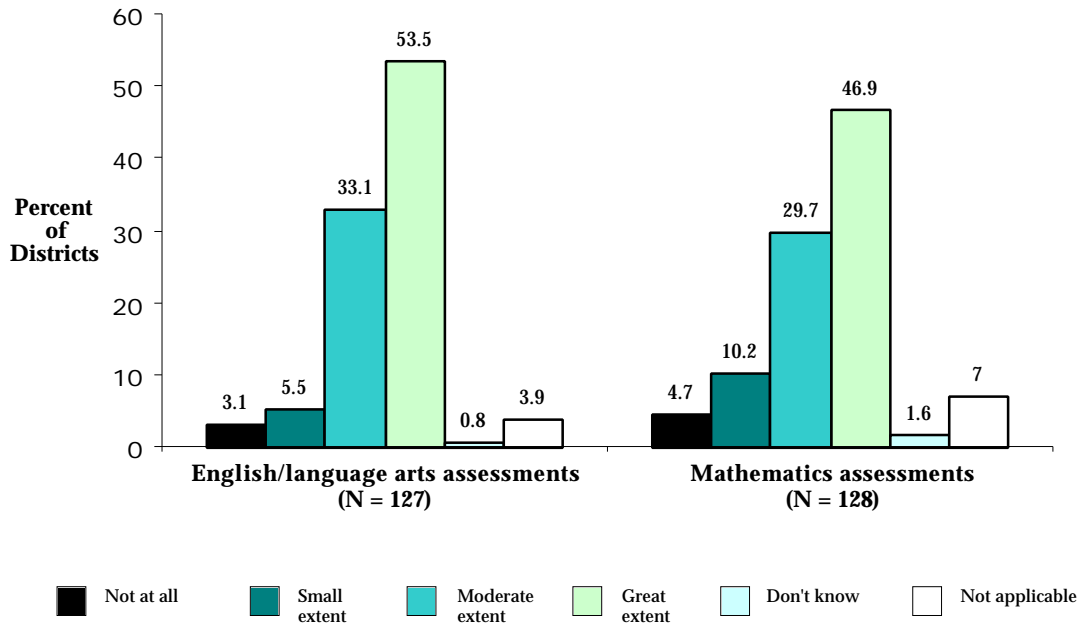
Ideally, local assessments should be aligned with content standards. To the extent that district assessments are used for accountability purposes and thereby help drive instruction, the use of standards-aligned assessments bodes well for standards-based instruction. Figure 7.1 shows that over 50 percent of survey respondents said that district English/language arts assessments are aligned with district content standards to a “great” extent (33.1 percent reported a “moderate” extent); for mathematics, 46.9 percent of respondents reported alignment to a “great extent” (with 29.7 percent citing a “moderate” extent).<sup>2</sup>

---

<sup>2</sup> District reporting about alignment may have been overly optimistic; these data have not been verified.

**Figure 7.1**

**District Reporting of the Extent to Which District Assessments are Aligned with District Content Standards for English/Language Arts and Mathematics**



For some districts, the development of standards and aligned assessments has been a parallel, or even an iterative, process. For example, one district in which researchers conducted interviews found that it had to reassess its standards after piloting district assessments, since several of the standards proved too difficult to measure.

### Use and Combination of Multiple Measures

**Many districts have developed multiple measures for every grade level, going beyond the state requirements.**

As discussed in Chapter 5, 62.4 percent of district survey respondents indicated that their districts had “fully implemented” the use of multiple measures to determine whether students meet grade-level standards. Similarly, the CDE review of 190 district accountability plans found that 69.5 percent of districts employed multiple measures in the required grade spans. More strikingly, only 6 percent did not include multiple measures in at least one grade level within the three IASA grade-level spans (grades 3–5, 6–9, and 10–12). The

independent review of 36 plans, meanwhile, found that 80.6 percent of districts employed multiple measures in the appropriate grades.

Many districts, moreover, appear to have gone beyond the requirements and have instituted multiple measures in all or most grade levels. The CDE-reviewed plans indicated that 38.4 percent of districts employed multiple measures at *all* grade levels, and separately commended about 25 percent of the plans where districts had “exceeded the minimum requirements and [had] multiple measures at all or most grades.” The independent plan review identified 58.3 percent of districts whose plans indicated multiple measures in all or most grades.

Two of the districts where interviews took place had also substantially exceeded the state multiple measures requirements, employing three measures. For one district (which was developing standards and multiple measures well before the state requirements were in place) the measures were the SAT-9, a district assessment, and teacher judgment. For the other district, the measures used were a standardized test, grades, and a writing sample scored with a standardized rubric.

**Districts appreciated the state guidelines on combining multiple measures, although some expressed frustration about the timing.**

Once districts have selected the measures they will use, they must figure out how to combine them to determine whether students meet minimum criteria for proficiency, and they must also decide what those criteria will be. Among the 190 CDE-rated plans, nearly 25 percent of the plans did not explain how the results from all measures were combined at the student level to determine if the student meets standards. Similarly, 7 of the 36 independently reviewed plans (19.4%) included no mention of how scores were combined to determine whether students meet standards.

As for districts that did discuss the combination of measures, 47.9 percent of the CDE reviewed plans (and 63.9 percent of the independently reviewed plans) used the CDE compensatory model.<sup>3</sup> The CDE conjunctive model, meanwhile, was used by 22.1 percent of the CDE plans (and only 5.6 percent of the independently reviewed plans). A weighted model was used by 13.2 percent of the CDE reviewed plans and 19.4 percent of the independently reviewed plans.

As with content standards, a number of districts had already started to develop a method of combining measures by the time the state came out with its guidelines. Some of these

---

<sup>3</sup> In the compensatory model, a student’s low performance on one measure can be offset by high performance on another. In the conjunctive model, students must score at or above certain absolute cut-off points on all measures to be considered as meeting grade-level standards (School and District Accountability Division, CDE, 1998, Spring).

districts thus had to revise their systems. This was discussed in several interviews. One district, for example, used in their first year of multiple measures a sum-of-points system for combining the measures. The following year, the state guidelines came out, and the district adopted the state's weighting tables for combining multiple measures, with some minor modifications. District staff were annoyed about the wasted work, but they acknowledged that having a standardized system for combining measures merited the switch.

Another interview district, however, was less pleased. They spent a year developing a rubric for combining multiple measures of assessment, only to receive later the CDE memo outlining acceptable practices. The district then had to redo its entire rubric.

## Assessment and the Public Schools Accountability Act

*I would hope we're being held accountable. The problem I see is that I don't think it's [the STAR test] the one way you test for that. I think it should be just one of a variety of things. But I definitely think we should be held accountable for student performance. If not, we're not doing our jobs....I just don't think it [accountability] should be measured with one set of tests, and that's it. The kids I have...are good kids; they came in with good scores, they'll go out with decent scores; they probably could have done that no matter whether I did a good job or not. On the other hand, you can get kids that are ill-prepared, and you know, how much you can help them improve — I don't know that anybody knows, is that 5 percentage points? Is that 25 percentage points? I guess we're all wondering, what's going to be the measure of achievement? So, that's all a little iffy when the test is the thing.*

—Teacher

**Districts are gravely concerned about the API's reliance on the SAT-9. They fear that as a result, the test — although not aligned with standards — will drive the entire education program.**

Although the Academic Performance Index (API) established by the Public Schools Accountability Act of 1999 eventually will be composed of several measures, at present the only measure for which reliable and valid data are available is the STAR test (currently the SAT-9). Thus, for the first two years, the API will be composed solely of these test scores.

As with the previous accountability requirements, districts raise many concerns about the weight given to the SAT-9 in the new legislation. One survey respondent wrote, "Major concerns over myopic use of standardized tests (SAT-9) for all major accountability decisions." Nearly every administrator who was interviewed expressed similar concerns. One, for example, said that her "absolute primary concern" with the PSAA was "the single

indicator” (i.e., the SAT-9); another said that there is “considerable concern” about the sole measure being the SAT-9.

A few administrators said that they worry that the new system will result in teachers teaching directly to the SAT-9. As one administrator put it, the new system may help students learn answers to the test questions as teachers do everything they can to improve test scores, but even if the test scores do rise, this may not be good for students’ overall education. (However, as indicated earlier in this chapter, teachers may already have been “teaching to the test” even prior to the passage of the PSAA.)

In another district, the Assistant to the Superintendent expressed regret that the district must now use the SAT-9 instead of the standardized test that they have been using, for which they have longitudinal results. Moreover, she commented, the SAT-9 is unrelated to their content standards.

The SAT-9’s lack of relationship to standards also came up several times in survey respondents’ written comments about the new legislation:

*The accountability (API) proposal may run counter to standards-based reform in its emphasis solely on STAR.*

*The idea of including information about attendance, days-outs and alternate assessments is very good, except that this data is currently unreliable or unavailable across the state. So, de facto we will all be forced to rely on SAT-9 which is inappropriate measure for accountability to a standard.*

*We would prefer to work with local standards and assessments which align tightly with our instructional program.*

Similarly, several districts’ survey comments pointed out that the SAT-9 is a norm-referenced test, and thus questioned its role in a standards-based system, suggesting that a criterion-referenced test might be more appropriate:

*SAT-9 is not a reasonable assessment for accountability. Assessments aligned to standards are critical. CRTs are necessary for measuring accountability for standards.*

*Would prefer to have criterion-reference test versus norm-referenced.*

*Norm-referenced test scores are most affected by family income level and the educational attainment of the parents. We cannot influence these elements. We need a fair, criterion-referenced testing system based on the standards.*

One district’s Director of Research and Evaluation mentioned in an interview that he is “somewhat concerned” about the measures of assessment that the state is developing. He supports the matrix assessment, and, in lieu of that test, supports using just the augmented sections of the SAT-9, since he feels the items on the basic-skills test are not aligned with the standards.

**Districts value multiple measures of student achievement and do not want to see them abandoned.**

Districts are even more concerned about the weight of the SAT-9 in the new system than in the previous system, because in the previous system, the use of multiple measures offset the emphasis on the SAT-9. As such, districts particularly object to the *increased* emphasis on the SAT-9 at the expense — or *loss* — of multiple measures. Indeed, as the following survey comments illustrate, districts strongly want to keep multiple measures, both because multiple measures are useful and because considerable effort went into their development:

*CA swinging back to dependence on standardized test only rather than multiple measures will be devastating.*

*If the API replaces multiple measures, I feel that will be unfortunate. The API as well as STAR testing doesn’t have the diagnostic benefit or staff buy-in of multiple measures.*

*It’s a shame to waste the efforts for implementing multiple measures. We will continue.*

*SBX\_1 is too narrow. Districts should be required to continue to report multiple measures as called for by IASA and the state’s plan for Title I.*

*The new X1...basically does away with multiple measures. The SAT-9 is a snapshot in time whereas multiple measures is accountability system for the entire year.*

*Concerned about loss of multiple measures as an enormous amount of time and energy was expended to do a terrific job and now will go out the window! It is hard to motivate teachers and to create one program that will be later replaced by AP Index from state. Why have us do it?*

*Just as we begin to assimilate multiple measures, the rules change and we start all over with an API index.*

Similar comments were made in district interviews. One administrator said that her district would like *more* multiple measures, not fewer; a second administrator in this same district said, “the more measures, the better,” and felt that “fairness” measures such as socioeconomic status and mobility should also be included in any accountability system. In

another district, whose development of multiple measures predated the state requirements, an administrator thought that the new API system represents a divergence from the district's reform efforts. She said, however, that the district would continue with the development of multiple measures of student achievement and standards-based assessment.<sup>4</sup> In a third district, administrators mentioned that they would prefer to retain multiple measures, but are afraid to get "too far ahead" of the state, fearing "yet another shift in direction." (See Chapter 11, "Challenges and Assistance," for further discussion of "shifts in direction.")

## In the Next Chapter

Bringing assessments into alignment with content standards is a necessary step in building an accountability system. Some districts perceived aligned assessments themselves as the endpoint, rather than as one step towards accountability. Many districts had embraced the practice of using multiple measures of student achievement for determining whether students had met grade-level standards, as this chapter has shown. However, it is necessary to analyze the data generated by multiple measures for districts to use assessments as a tool for continuous improvement. The use of data is discussed in the next chapter.

---

<sup>4</sup> A different administrator in this district expressed a concern that the 40 percent of API index will be based on non-academic factors such as attendance and graduation rates. He feels that these are factors that schools are largely unable to influence, so it is not fair for them to count as such a large part of the index. Other people, however, did not seem to share this concern, as reflected by comments made on the survey and in interviews.