The results are in, and they are concerning: Study after study confirms that students lost critical opportunities to learn during the coronavirus disease 2019 (COVID-19) pandemic, and historically disadvantaged students bore the brunt of these lost learning opportunities (e.g., Goldhaber et al., 2022; Halloran et al., 2021; Kuhfeld, Soland, and Lewis, 2022; Nation’s Report Card, undated). As the COVID-19 pandemic continues to recede, schools’ attention has turned to helping students recover from several years of disruptions to schooling.

As emphasized by the U.S. Department of Education, assessment systems are key policy tools to identify students’ learning needs, target needed supports, and ultimately monitor student progress toward recovery (U.S. Department of Education, 2021). Schools administer many types of assessments each school year—all of which have different purposes and stakeholders. For instance, assessments include low-stakes, small-scale formative assessments that are embedded within classroom instruction and typically rapidly administered. Such assessments help teachers in identifying and understanding students’ learning gaps, thereby allowing them to improve their instruction and student learning. On the other end of the spectrum, schools also administer summative assessments, which are assessments designed to assess student learning against a set of learning objectives or standards at the end of a unit, course, semester, or school year, for such purposes as evaluation or accountability. High-stakes end-of-year statewide assessments that are tied to federal accountability requirements fall within this latter category (Dadey and Diggs, 2019; Ho, 2022; Perie, Marion, and Gong, 2009).

In this report, we draw on the framework developed by Perie, Marion, and Gong, 2009, to focus on benchmark assessments—a type of assessment that prior literature has situated somewhere between formative and summative assessments. Benchmark, or interim, assessments are defined as assessments which “evaluate students’ knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school, or district level” (Perie, Marion, and Gong, 2009). Because benchmark assessments occupy this middle
Benchmark assessments are also predictive in that they are meant to provide information on students’ expected performance on summative assessments. Yet, similar to formative assessments, benchmark assessments also have an instructional role in that they can be used to identify students’ strengths and areas for growth and inform teachers’ instructional decisions. Thus, benchmark assessments may have different uses for different stakeholders. Benchmark data aggregated at the school or district level for evaluative or predictive purposes may be relevant to leaders at those levels, although benchmark data at the classroom level that are used to identify and address student learning gaps may be most relevant to classroom teachers.

In the wake of heightened federal accountability pressures following the enactment of No Child Left Behind in 2001, many districts adopted benchmark assessments to gather more information on students’

---

**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIRS</td>
<td>American Instructional Resources Survey</td>
</tr>
<tr>
<td>ASLP</td>
<td>American School Leader Panel</td>
</tr>
<tr>
<td>ATP</td>
<td>American Teacher Panel</td>
</tr>
<tr>
<td>COVID-19</td>
<td>coronavirus disease 2019</td>
</tr>
<tr>
<td>ELA</td>
<td>English Language Arts</td>
</tr>
<tr>
<td>MAP</td>
<td>Measures of Academic Progress</td>
</tr>
</tbody>
</table>
progress toward meeting proficiency targets on statewide summative assessments (Dadey and Diggs, 2019), spurring an increased emphasis on data-driven instruction (Datnow and Hubbard, 2015). As schools continue to redress interruptions to student learning brought on by the COVID-19 pandemic, schools’ and teachers’ effective use of assessment data to advance student learning takes on a new urgency. We might expect benchmark assessment data to play a critical role in this effort, given limitations of other assessment types. For example, summative assessments have been criticized for providing data too late in the school year to meaningfully inform instruction. Meanwhile, data from formative assessments are not designed to be aggregated up beyond the classroom level and thus are not well suited to make comparisons in student achievement within and across schools. However, benchmark assessments themselves also are not immune to criticisms. For instance, prior literature has noted the limited evidence linking use of benchmark assessments to improved student achievement, limited understanding of how benchmark assessments might lead to those improvements, and concerns about the grain size or timing of the resulting data (Dadey and Diggs, 2019; Datnow and Hubbard, 2015; Konstantopoulos et al., 2016; Marion, 2018).

Importantly, assessment systems (and benchmark assessments specifically) do not exist in isolation. Benchmark assessments are only one element of a coherent instructional system. In a coherent instructional system, all elements of the instructional system—including benchmark assessments, summative assessments, curriculum materials, academic standards, professional development, and teacher evaluation systems—are aligned such that the elements work together cohesively to fulfill the varied needs of different stakeholder groups (Wang et al., 2022; Polikoff et al., 2020). As school systems emerge from the chaos wrought by the COVID-19 pandemic and opportunities for gathering student assessment data resume, education leaders and policymakers have an opportunity to take stock of the state of their assessment systems and reflect on whether assessments—and benchmark assessments specifically—are appropriately aligned with other aspects of schools’ instructional systems.

**Assessment systems do not exist in isolation. Benchmark assessments are only one element of a coherent instructional system.**

**Our Current Study**

Despite heightened emphasis on data-driven instruction and widespread use of benchmark assessments, we do not know much about what kinds of benchmark assessments schools are using. Nor do we know how the prevalence of various types of benchmark assessments has shifted over time. And despite literature highlighting the importance of connections between benchmark assessments and other aspects of instructional systems—such as curriculum, professional learning, summative assessments, and academic standards (Herman, 2017; Marion, 2018; Perie, Marion, and Gong, 2009)—we know little about educators’ perceptions of how benchmark assessments fit in with these other aspects of their schools’ instructional systems. This report fills in some of these key information gaps, focusing on two key research questions:

1. How prevalent were benchmark assessments in the 2021–2022 school year, and how has the prevalence of different types of benchmark assessments changed over time?
2. To what extent do educators perceive their benchmark assessments to be aligned with other aspects of their schools’ instructional systems?

To answer these questions, we primarily rely on data from the American Instructional Resources Survey (AIRS). AIRS was administered to nationally
Nearly all K–12 public schools (99 percent) said they administered at least one benchmark assessment to at least some of the students in their school during the 2021–2022 school year. Representative samples of teachers and principals via RAND’s American Educator Panels in four consecutive school years (2018–2019 to 2021–2022). The survey included items asking educators about their use of various benchmark assessments and whether they perceived benchmark assessments as aligned with other aspects of their instructional system. We supplement these survey data with data from 45 interviews conducted with K–12 teachers. (For more information about the data sources used in this report, see Box 1.)

We present our results in two sections. In the first section, we use data from principals to examine the widespread use of benchmark assessments in the most recent school year (2021–2022) and show how schools’ use of specific benchmark assessments has grown over time. To generate these prevalence estimates, we leverage principals’ survey responses because schools or districts—rather than teachers—are often in the position of selecting which benchmark assessments to administer in their schools (Datnow and Hubbard, 2015; Herman, 2017).

In the second section, we rely on survey and interview data from teachers and principals to investigate educators’ perceptions of benchmark assessments’ alignment (and misalignment) with other key elements of coherent instructional systems. We hypothesize that educators’ perceptions of coherence might shed light, in part, on the extent to which the benchmark assessments currently in use are able to fulfill their evaluative, predictive, and instructional roles. We use principals’ reports about their perceptions about alignment between benchmark assessments and summative assessments and academic standards, given the importance that such alignment may have for benchmark assessments’ evaluative and predictive uses. In contrast, teachers are in the position of making sense of how to use assessment data to drive their classroom instruction. Therefore, when discussing teachers’ perceptions of coherence with regards to benchmark assessments, we focus explicitly on benchmark assessments’ instructional purpose. We zero in on teachers’ perceptions of alignment between their benchmark assessments and their curriculum materials as well as their access to professional learning experiences which support their use of assessment data.

The Prevalence of Benchmark Assessments

In 2021–2022, there was widespread use of benchmark assessments. In K–8 schools, published benchmark assessments were somewhat more common than locally created assessments in 2021–2022. In high schools, it was the opposite.

Nearly all K–12 public schools (99 percent) said they administered at least one benchmark assessment to at least some of the students in their school during the 2021–2022 school year. Schools reported administering a wide variety of benchmark assessments in the 2021–2022 school year. Throughout the remainder of this report, we discuss what we refer to as locally created benchmark assessments, by which we mean assessments created by state departments of education, school districts, or individual schools. We contrast these assessments with published benchmark assessments, or those created by external vendors that schools purchase off the shelf.

We consider to what extent schools’ use of locally created and common published benchmark assessments in the 2021–2022 school year varied by school grade level. We use K–8 schools to refer to schools
Box 1. Description of Data Sources

We leverage three sources of data—principal survey data, teacher survey data, and teacher interview data—to conduct the analyses presented in this report. We briefly describe each data source in this box.

Principal survey data: To generate national prevalence estimates for various types of benchmark estimates, we used principal data from AIRS. AIRS was administered to nationally representative samples of roughly 1,500 principals in the spring of four consecutive school years (from 2018–2019 to 2021–2022) via the RAND Corporation’s American School Leader Panel.

Each school year, principals were asked what benchmark assessments their schools administered in ELA and math in that year to assess student learning. The list of benchmark assessments included both generic options for locally created assessments (i.e., state-created, district-created, school-created assessments) in addition to options for common published benchmark assessments (e.g., i-Ready, Measures of Academic Progress [MAP] Growth). A complete list of the benchmark assessments asked about on each principal survey are included in Table A.1 in the appendix.

Principals were asked separately about benchmark assessments for ELA and math and were not asked about use of benchmark assessments in other subjects. We initially examined patterns in principal reports of use of benchmark assessments separately for ELA and math. However, because results were nearly identical, we present prevalence estimates using only data on ELA assessments for brevity. Because results are so similar, we avoid labeling results as pertaining to “ELA benchmark assessments” only and instead interpret the prevalence estimates as applicable to benchmark assessments in both ELA and math.

Furthermore, we interpret principals’ responses to these survey items as speaking on behalf of their school. Therefore, we report prevalence estimates using the school as the unit of analysis (e.g., “percent of schools who administered benchmark assessment”). However, we note that a handful of our survey respondents (less than 1 percent) are matched to the same school, so this terminology should be interpreted with some caution.

Teacher survey data: To present selected findings on teachers’ perspectives about their benchmark assessments, we use data from the 2021–2022 AIRS teacher survey. This survey was administered to a nationally representative sample of K–12 ELA, math, and science public school teachers through the RAND Corporation’s American Teacher Panel (ATP) in April and May 2022. We focused on the responses of ELA and math teachers ($n = 6,399$).

Teacher interview data: In May and June 2022, a team of five qualitative researchers and analysts conducted 45 interviews with K–12 teachers. Interviewees were drawn from a nationally representative sample of ELA and math teachers who were part of the RAND ATP in spring 2022 and who completed a survey from a related project on coherent instructional systems. In the interviews, we asked teachers about the alignment between their instructional system components (e.g., curriculum materials, academic standards, professional development, teacher collaboration, evaluation, summative assessments, and benchmark assessments). Our analysis further focused on excerpts in which teachers explicitly expressed perceptions about their benchmark assessments, unless otherwise noted. Because a focus on benchmark assessments was just one small part of the overall interview protocol and we draw on the responses of a modest sample of teachers, teachers’ responses are meant to be illustrative of survey findings and not generalizable to the perspectives of all teachers. To learn more about this interview study, see Pauketat et al., 2023.

serving any grades kindergarten through grade 8 and high schools to refer to schools serving any grades 9 through 12. Importantly, these prevalence estimates capture only whether schools administered particular assessments at all in the 2021–2022 school year and not specific information around assessment conditions (e.g., which and how many students were tested, how many times, for what reason). For example, although a high school may have administered a specific published benchmark assessment in the 2021–2022 school
year, they may have administered the assessment only to students in specific grades, in specific courses, or who met certain conditions.

With this caveat in mind, we find that locally created assessments were widely used in K–8 schools and high schools in the 2021–2022 school year, although these assessments were slightly more common in high schools. Eighty-four percent of high schools reported administering at least one locally created assessment in the 2021–2022 school year, as did 77 percent of K–8 schools (see Figure 1). Among K–8 schools, state- and district-created assessments were more prevalent than school-created assessments: Roughly half of K–8 schools administered a state- or district-created assessment, and only one-third administered a school-created assessment.

Both K–8 schools and high schools also reported widespread use of published benchmark assessments in the 2021–2022 school year, although the use of published benchmark assessments was somewhat more common in K–8 schools than in high schools. Eighty-seven percent of K–8 schools administered a published benchmark assessment to at least some students in the 2021–2022 school year, compared with 62 percent of high schools. More K–8 schools administered published benchmark assessments than administered locally created assessments, but published benchmark assessments were less common in high schools than locally created assessments.

There are many published benchmark assessments on the market, and the principals who took our survey indicated that their schools used a wide variety of the roughly 30 published benchmark assessments we asked about for the 2021–2022 school year. Because it is infeasible to discuss findings for all published benchmark assessments, we focus on use of three specific published assessments: i-Ready (produced by Curriculum Associates),2 Star (produced by Renaissance Learning),3 and MAP Growth (produced by NWEA). We chose these assessments to illustrate pat-

---

**FIGURE 1**

Percentage of Schools Reporting That Their Students Took Various Types of Benchmark Assessments in 2021–2022, by School Grade Level and Assessment Type

<table>
<thead>
<tr>
<th></th>
<th>Elementary/Middle School (K–8)</th>
<th>High School (9–12)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Locally created</td>
<td>Published</td>
</tr>
<tr>
<td>Any published</td>
<td>77</td>
<td>29</td>
</tr>
<tr>
<td>State-created</td>
<td>50</td>
<td>24</td>
</tr>
<tr>
<td>District-created</td>
<td>50</td>
<td>22</td>
</tr>
<tr>
<td>School-created</td>
<td>35</td>
<td>22</td>
</tr>
</tbody>
</table>

NOTE: This figure depicts response data from the following survey question: “Which benchmark assessments do your students take over the course of this school year (2021–22) to assess their progress in ELA and mathematics?” (n = 1,364) Principals were asked separately about benchmark assessments in ELA and math. Because results were similar, we present only results for ELA benchmark assessments for brevity. Our survey asked principals separately about “i-Ready Assessments,” “i-Ready Diagnostic,” and “i-Ready Standards Mastery.” We use data from these three subitems to determine which schools used “any i-Ready” assessment. Similarly, our survey asked principals separately about “Star Assessments (Renaissance)” and “Star Reading/Star Math (Renaissance Learning).” We use data from these two subitems to determine which schools used “any Star” assessment. Black bars represent 95-percent confidence intervals.
terns for two reasons. First, all three assessments can be administered to students in any grades K–12 (Curriculum Associates, undated; Renaissance Learning, undated; Shields Fleming, 2021), allowing us to make prevalence comparisons across school grade levels. Second, according to principals’ reports, i-Ready, Star, and MAP Growth were the three most prevalent assessments administered in K–8 schools and were among the top four in high schools as of the 2021–2022 school year. (To see what proportion of schools administered other published benchmark assessments, see Table A.1 in the appendix.)

According to principal reports, roughly three in ten K–8 schools administered i-Ready to at least some students in the 2021–2022 school year, and roughly one-quarter of K–8 schools each administered MAP Growth or Star. Meanwhile, two in ten high school principals or fewer reported that their school administered MAP Growth, Star, and i-Ready to some students in any grades 9 through 12.

According to principals’ reports, administration of multiple different benchmark assessments, including those from different providers, in the same subject area was common in the 2021–2022 school year. For example, a high school might administer a state-created benchmark assessment, a school-created benchmark assessment, and a published benchmark assessment in the same subject (although all assessments might not be administered to all students). Schools administered three different benchmark assessments per subject, on average, in the 2021–2022 school year. A small minority of schools (15 percent) administered only one benchmark assessment per subject in the 2021–2022 school year.

Schools may choose to administer multiple benchmark assessments in a single subject in a single school year because different assessments might be more appropriate for specific school or teacher use cases. For example, a high school’s leadership may choose to administer a school-created assessment because they believe it is closely aligned with teachers’ daily teaching and, therefore, that data from the assessment might help teachers identify how to adjust their instruction to help struggling students. The same school might also choose to administer a published benchmark assessment in the same subject to monitor students’ progress toward meeting proficiency standards in preparation for state summative assessments.

We generally did not observe meaningful significant differences in the prevalence of locally created or published assessments in the 2021–2022 school year by school context or by principals’ perceptions of their school’s historic achievement level, with a few exceptions. According to principal reports, district-created assessments were more prevalent in the 2021–2022 school year in urban and suburban schools, in schools with larger enrollment sizes (those serving 450 or more students), and in schools serving mostly students of color. We interpret these patterns as evidence that district-created assessments were more common in larger school districts in the 2021–2022 school year. We hypothesize that these district-created assessments may be more prevalent in these settings, both because these districts have the resources to produce assessments and because it may be cost-effective to do so.

Use of published assessments has grown steadily but marginally over the past four school years. Schools’ use of locally created assessments also grew, although there were large swings in 2019–2020 and 2020–2021, presumably because of pandemic-related disruptions to schooling.

We now turn to consider how schools’ use of specific benchmark assessments has grown over time. Schools’ use of published benchmark assessments has grown steadily, though marginally, over the past four school years. Between the 2018–2019 and 2021–2022 school years, there have been 7 percentage-point and 10 percentage-point increases, respectively, in the share of K–8 schools and high schools administering at least one published benchmark assessment to least some students (see Figure 2). These overall increases were driven by modest increases in the share of schools administering i-Ready and MAP Growth over this period. For example, use of i-Ready grew by 11 percentage points over the past four school years in K–8 schools and grew by 7 percentage points in high schools, at least according to principal reports.
We observed a different pattern in the prevalence of locally created assessments over the past four school years. For all three types of assessments, we observed a U-shaped pattern: an increase in the share of schools administering these assessments between 2018–2019 and 2019–2020, a decline in 2020–2021, and then an increase again in the 2021–2022 school year. We hypothesize that this U-shaped trend stems from pandemic-related disruptions to schooling. Specifically, schools’ use of locally created bench-
mark assessments may have been unusually high in 2019–2020 because they scrambled to obtain some measure of how much students learned throughout the school year in the absence of state summative assessment data. Or, because schools were excused from the burden of end-of-year summative assessments and federal accountability pressures (Gewertz, 2020), they may have pivoted to administering assessments they perceived as having more utility, such as those designed internally to more directly align with teachers’ instructional practices (Braun and Marion, 2022).

We observed a decline in the proportion of schools administering locally created assessments between 2019–2020 and 2020–2021, possibly explained by the abnormality in schooling conditions that persisted throughout 2020–2021 and the re-introduction of state summative assessments (Institute of Education Sciences, undated; Gewertz, 2021). However, the share of schools using various locally created assessments in 2020–2021 was roughly on par with—or even higher—than in 2018–2019. The fact that schools’ use of locally created assessments in 2020–2021 was at least on par with their prepandemic use suggests that perhaps 2020–2021 appears as a setback only because of abnormally high reliance on these assessments in 2019–2020.

Although we observed the same U-shaped growth pattern for all three types of locally created assessments, schools’ use of each assessment type did not grow at the same rate, particularly in K–8 schools. For example, the share of K–8 schools administering district-created assessments increased more over this period than the share using school-created assessments (20 percentage–point growth versus 13 percentage–point growth, respectively).

In summary, we interpret these trend data as an indication that, despite swings, more schools were relying on locally created assessments in the 2021–2022 school year than in the 2018–2019 school year. We hypothesize this may be for several reasons. For one, locally created benchmark assessments have become increasingly important for schools and teachers looking to make data-informed decisions—a trend that began before the COVID-19 pandemic started. Second, schools likely are using these assessments to obtain real-time information on students’ progress as they recover from decreased opportuni-

ties to learn during the pandemic (Nation’s Report Card, undated).

Most importantly, these data suggest that benchmark assessments—both published and locally created—were already widely used before the pandemic began, and the share of schools administering different types of benchmark assessments has only increased over the past four school years. Thus, widespread use of these assessments suggests that we should be concerned about how these assessments are fitting into schools’ broader instructional systems.

The Role of Benchmark Assessments in Coherent Instructional Systems

In this section, we consider the extent to which educators perceive benchmark assessments to be aligned with other components of their instructional systems, such as curriculum, standards, and summative assessments. We interpret perceptions of alignment as an indication of instructional system coherence. To learn more about what we mean by instructional system coherence and how we define various elements within instructional systems, see Box 2.
which are critical for evaluating benchmark assessments’ evaluative and predictive purposes. Regardless, Perie, Marion, and Gong, 2009, notes that, with regards to benchmark assessments’ evaluative role, it is critical that the assessments are “targeted to the content standards that are the focus of the educational program(s) being evaluated or studied” (p. 10). Similarly, with regards to benchmark assessments’ predictive role, benchmark assessments should be aligned to summative assessments if they are to predict students’ performance on those assessments.

Regardless of what specific benchmark assessment(s) their school administered, principals overwhelmingly perceived these assessments to be mostly or totally aligned with the content of their state standards and state summative assessments. For example, 89 percent of principals whose schools administered district-created benchmark assessments perceived them to be aligned with their state standards, and 87 percent said district-created bench-

More than 80 percent of principals perceived their schools’ benchmark assessments as aligned with their state standards and state summative assessments

We asked principals who indicated that their school had administered benchmark assessments in the 2021–2022 school year whether they perceived these assessments to be aligned with two other key elements of their schools’ instructional system: state standards and summative assessments. Principals’ perceptions about alignment between their school’s benchmark assessments on one hand and state standards and summative assessments on the other in particular can help us better understand the extent to which principals believe benchmark assessments are fulfilling their evaluative and predictive roles, although we acknowledge there are many other factors described by Perie, Marion, and Gong, 2009, which are critical for evaluating benchmark assessments’ evaluative and predictive purposes. Regardless, Perie, Marion, and Gong, 2009, notes that, with regards to benchmark assessments’ evaluative role, it is critical that the assessments are “targeted to the content standards that are the focus of the educational program(s) being evaluated or studied” (p. 10). Similarly, with regards to benchmark assessments’ predictive role, benchmark assessments should be aligned to summative assessments if they are to predict students’ performance on those assessments.

Regardless of what specific benchmark assessment(s) their school administered, principals overwhelmingly perceived these assessments to be mostly or totally aligned with the content of their state standards and state summative assessments. For example, 89 percent of principals whose schools administered district-created benchmark assessments perceived them to be aligned with their state standards, and 87 percent said district-created bench-

---

**Box 2. Definition of Instructional System Components**

When describing the instructional systems in which educators work, we refer to various instructional system components. Beyond benchmark assessments, which we defined at the beginning of this report, we focus on the following instructional system components:

- **Standards** outline the concepts, knowledge, and skills, often codified at the state level, that students are expected to master at grade or course level within a content area.
- **Curriculum materials** are instructional materials (e.g., specific textbooks or instructional programs) adopted for use by teachers and/or their school or district. Curriculum materials are intended to constitute a full course of study for a subject and grade level.
- **State summative assessments** are the high-stakes, state-required assessments given to students in select grades and subjects in the spring of each school year to evaluate student learning against state standards and fulfill states’ standardized testing requirements under the Every Student Succeeds Act.
- **Professional development** consists of formal structures presented by schools and districts to support or improve an aspect of instruction through the growth of teachers.
- **Teacher collaboration** refers to opportunities for teachers to work with peers to make meaning for their instruction. This includes professional learning communities or common planning time.
- **Teacher preparation** refers to teachers’ experiences that provided them with their initial credential to teach in a K–12 school. These include teachers’ program coursework and clinical teaching experiences, where they taught in a classroom under the supervision of a mentor teacher.

**Coherence, or alignment**, is the extent to which these instructional system components are consistent with each other in providing the same signals and supports to teachers and leaders about what instruction should look like. Conversely, **incoherence (misalignment)** occurs when one or more components are not linked to the larger instructional system and/or provide conflicting signals to teachers about what instruction should look like.

*The definitions presented in this box are largely drawn from Pauketat et al., 2023.*
marks were mostly or totally aligned with the content of their state summative assessment. We observed the same high level of agreement for published benchmark assessments. For example, 90 percent of principals whose schools administered i-Ready agreed that assessments were mostly or totally aligned with state standards, and 88 percent answered similarly about summative assessments. In sum, principals’ agreement about benchmark assessment alignment was high—above 80 percent—for all assessments and for alignment with both standards and summative assessments. Principals’ perceptions about alignment between their benchmark assessments, on one hand, and summative assessment and standards, on the other, also held high across grade bands (i.e., K–8 and high schools) and subjects (i.e., ELA and math) with little variation. The fact that the vast majority of principals perceived alignment is perhaps unsurprising because school systems have autonomy in determining what benchmark assessments to administer and to what grade levels. Thus, schools may be choosing to administer the benchmark assessments that they perceive to have the greatest alignment with their state standards and assessments. However, we only asked about principals’ perceptions of alignment for their ELA benchmark assessments and for their math benchmark assessments rather than asking about each individual benchmark assessment that their school used; thus, it is likely that principals were aggregating their perceptions of alignment across grade levels and assessment types, especially given that our findings suggest that schools typically use multiple benchmark assessments per subject.

Two-thirds of teachers reported their benchmark assessments were aligned with their curriculum materials, but teachers’ perceptions of alignment depended upon various factors, such as their subject, the type of assessments they used, the students they taught, and their professional learning experiences.

In this section, we focus explicitly on benchmark assessments’ instructional purpose. We explore teachers’ perceptions of alignment between benchmark assessments and their curriculum materials and examine how their perceptions of alignment may be influenced by various factors, including the materials and benchmark assessments they use; the students they serve; and other instructional system components, such as their professional learning and teacher preparation experiences. Teachers are particularly well suited to speak to the alignment of these aspects because they are tasked with using student assessment data to guide their instruction.

Furthermore, we examine why alignment between benchmark assessments and curriculum materials matters by investigating the relationship between teachers’ perceptions of alignment and their ability to effectively use student data to meet students’ needs.

To better understand teachers’ perceptions about how benchmark assessments align with other instructional system components, we asked teachers the extent to which the benchmark assessments their students take align with the content of the curriculum materials they use regularly for their ELA or math instruction. This connection is especially important to explore because one role of benchmark assessments is to assess students’ performance on the standards teachers are teaching (Perie, Marion, and

[Our] district created interim testing which is specifically aligned to the curriculum, so it’s meant to kind of mimic [the curriculum] to a pretty high degree as far as the questioning and types of problems that students are facing.

— Elementary math teacher
Vocabulary or background knowledge necessary to support comprehension, might be considered more aligned to teachers’ curriculum materials than cold reads, in which students have little such exposure. Furthermore, we hypothesize that students may be less likely to encounter warm reads unless assessments are designed to align with curricula.

Next, we examined whether teachers’ perceptions of alignment might be influenced by the type of curriculum material they used—specifically, whether they used a standards-aligned curriculum. To conduct this analysis, we leveraged a definition of standards alignment based on curriculum reviews by EdReports, an independent organization that reviews published curriculum materials and provides ratings based on whether they align with most college- and career-ready academic standards across the United States. We defined curriculum materials as standards-aligned if they are rated as fully meeting the expectations of college- and career-ready standards. One reason we might expect to see standards-aligned materials as supporting greater coherence between curriculum materials and benchmark assessments is that standards-aligned curriculum materials are rated as appropriately aligned to academic standards, and benchmark assessments ideally should also be aligned to academic standards. However, when we compared the responses of teachers who regularly or intensively used at least one standards-aligned material with those of teachers who did not regularly or intensively use any standards-aligned materials, we observed no difference in teachers’ perceptions about coherence between their benchmark assessments and curriculum materials for either ELA or math teachers.

Instead, our data suggest that the type of benchmark assessment that teachers used was linked to their perceptions about alignment. Earlier, we discussed principals’ reports of the rising prevalence of both locally created (e.g., state-, district-, or school-created) benchmark assessments and published benchmark assessments. Similarly, in the 2021–2022 school year teacher survey, we asked teachers to report on their use of benchmark assessments. One key difference in how we asked teachers about benchmark assessments, relative to principals, was that teachers were asked only whether they used “bench-
Box 3. Alignment and Misalignment Between Benchmark Assessments and Curricula

During our interviews, we asked teachers about whether they perceived their assessments and curriculum materials as aligned or misaligned. Notably, more teachers mentioned misalignment between their benchmark assessments and curriculum materials than mentioned alignment, even though our survey data suggest that the majority of ELA and math teachers nationally perceived their benchmark assessments as mostly or totally aligned with their curriculum materials. Nevertheless, through our interviews, teachers provided a glimpse of what alignment and misalignment look like in practice.

Alignment

Teachers perceived alignment when they felt that assessments matched their curriculum materials and lessons in content and format. For instance, one teacher who perceived alignment between their curriculum and benchmark assessments said, “It aligns because . . . if we’re covering . . . how to write a historical fiction piece . . . our district assessment will also be on historical fiction and writing a piece.” When describing alignment in terms of format, another teacher explained that their benchmark assessment was designed to match their curriculum and was one where “students have seen the problems before, just with different numbers.”

When speaking about alignment between their benchmark assessments and curriculum materials, teachers also described the following scenarios:

- About one-tenth of interviewed teachers described instances in which their districts or instructional coaches intentionally created interim assessments that were designed to match to their curriculum materials.
- About one-tenth of interviewed teachers mentioned that one manifestation of coherence was alignment across results from different assessment types, including their curriculum-embedded assessments, benchmark assessments, and summative assessment.
- About one-tenth of interviewed teachers viewed curriculum-embedded assessments favorably because they were “built into the program” and thus aligned with what students were doing in the classroom. As one teacher said, “The curriculum-based assessments are a lot better in terms of assessing what kids are actually capable of doing versus the state [practice] assessments that are like, ‘We have to give these practice tests to get them ready.’ . . . We get better data from the actual curriculum because it’s what they’re actually learning.”

Misalignment

Teachers who perceived misalignment between their curriculum materials and benchmark assessments felt there was a disconnect between what students are being assessed on and what they are actually learning in the classroom.

Teachers mentioned several different dimensions to this disconnect:

- **Content**: About 20 percent of interviewed teachers described scenarios in which their benchmark assessments tested students on concepts, skills, or topics which they had not taught in class. For instance, teachers described scenarios in which they were teaching to their state academic standards using their curriculum materials, but their benchmark assessments did not test those same standards. As one teacher put it, “We’re not testing what we’re teaching or being told to teach.”
- **Rigor**: About 20 percent of interviewed teachers expressed misalignment between the level of difficulty or expectations of their classroom instruction and the rigor of the assessments. Most of these teachers mentioned this disconnect with computer adaptive tests, such as i-Ready or MAP Growth, because students might be assessed on above-grade-level concepts that are not outlined in teachers’ standards or that teachers have not yet taught, even though this is how computer adaptive tests are designed to evaluate students’ knowledge.
- **Format**: About one-tenth of interviewed teachers expressed misalignment between how skills or concepts are presented or assessed in class and how they are presented on the assessment (e.g., differences in question wording or question format).
- **Timing**: Finally, three teachers also expressed misalignment in the timing of when students learned material in class and when they would be exposed to concepts on assessments or a lack of time to address the concepts on benchmark assessments. As this teacher said, “[We’re] assessing their fluency, but they don’t know those numbers, because they haven’t even been taught them yet, so it just doesn’t align with the calendar aspect of our curriculum.”
ELA and math teachers used benchmark assessments for their subjects they created themselves, and about 80 percent used published benchmark assessments required by their school or district. Notably, about 80 percent of teachers who used self-created benchmark assessments also used published benchmark assessments, although only about one-third of teachers who used published benchmark assessments also used self-created benchmark assessments.

Teachers using self-created benchmark assessments were more likely than teachers not using self-created benchmark assessments to perceive that their benchmark assessments mostly or totally aligned with their curriculum materials. This held true for both ELA and math teachers. For instance, 68 percent of ELA teachers using a self-created benchmark assessment felt that their assessments aligned with their curricula, in comparison with 57 percent of ELA teachers not using a self-created benchmark assessment and 60 percent of ELA teachers using a published benchmark assessment (Figure 3). Perhaps this is unsurprising because teachers creating their own benchmark assessments are able to engender a stronger connection between their benchmark assessments and curricular content because they are able to create questions or select topics that match what they’re teaching in class.

This hypothesis is also supported by the responses of about 20 percent of the teachers we interviewed, who created their own assessments to better align their lesson content to how they were assessing students. For instance, one teacher, expressing a lack of alignment between their state summative assessment and their curriculum, described creating their own assessments with other teachers in the school. The teacher said,

> I know what’s expected on the state tests, and I always feel like it’s a higher rigor than the students can perform at. So I feel like that also conflicts with even the way that we teach [the curriculum]... So that’s why we’re trying to develop as much as we can to bridge the gap there [by]... creating our own assessments.

Another teacher described a different scenario in which they perceived the district’s interim assessments as less rigorous than their curriculum...
Teachers’ Perceptions of Alignment Between Benchmark Assessments and Curriculum Materials, by Subject and Type of Benchmark Assessment

**FIGURE 3**

Teachers serving students performing below grade level perceived less alignment than their peers serving students at or above grade level.

Next, we explored whether teachers’ perceptions about the characteristics of their students might influence their perceptions of alignment. Teachers serving students they perceived as performing, on average, below grade level were less likely than their peers serving students at or above grade level to report alignment between their curriculum materials and benchmark assessments. This finding is concerning because it is plausibly even more important that teachers serving struggling learners perceive a connection between their curricula and assessments so that they can make data-driven decisions when adjusting their instruction to meet students’ needs.

"It’s just more of like an unreasonable level of expectation with the test. Whereas the curriculum on a daily basis what’s being asked of the kids is fair and reasonable, but what was asked of them on their last test . . . was just unfair and unreasonable." — Secondary math teacher

### NOTE

This figure depicts response data from the following survey question: “To what extent do the [ELA/mathematics] benchmark assessments your students take align with the content of the curriculum materials you regularly use for your [ELA/math] instruction?” To determine which type of benchmark assessment teachers used for their subject, we used teachers’ responses to a question about which benchmark assessments their students had already taken in the 2021–2022 school year to assess their progress in ELA and math. \( N = 5,653 \) [all ELA and math teachers]; \( n = 818 \) [math teachers using a self-created benchmark assessment]; \( n = 2,001 \) [math teachers using a published benchmark assessment]; \( n = 1,158 \) [ELA teachers using a self-created benchmark assessment]; \( n = 2,948 \) [ELA teachers using a published benchmark assessment]. Black bars represent 95-percent confidence intervals.
"We’re taught to differentiate instruction, but we give standardized tests.” Even with computer adaptive assessments, such as i-Ready or MAP Growth, where assessments are tailored to students, a few teachers expressed perceptions of misalignment because they felt that students were tested on concepts that they had not yet mastered or to which they had not been exposed. One teacher, expressing frustration, said, “Why are you testing the kids on this? … And especially in our class, we’re still struggling with regrouping. We’re not even on multiplication. That’s not even a second-grade standard.”

Teachers’ professional learning is linked to their perceptions of alignment between their benchmark assessments and curriculum materials

We hypothesize that teacher preparation and professional learning, including professional development training and peer collaborative learning, could act both as potential avenues to help teachers connect their benchmark assessments and curriculum materials and to make data-driven instructional decisions with assessment data, thereby bolstering teachers’
Perhaps unsurprisingly, our findings suggest that it is teachers’ ongoing professional learning—rather than their preparation experiences—that help them better connect their benchmark assessments and curriculum materials. Teachers that spent over one-quarter of their professional development or peer collaborative time on the use or analysis of student assessments were more likely than their counterparts who spent less time to report alignment between their benchmark assessments and curriculum materials (Figure 6). Although we observed the same patterns when we examined teachers’ responses according to how much time they spent in their preparation coursework or clinical experiences on using or analyzing student assessments,

**FIGURE 5**
Percentage of Teachers Reporting the Proportion of Time Spent During Their Professional Learning or Teacher Preparation Experiences Focused on Use or Analysis of Student Assessments

![Bar chart showing the percentage of teachers reporting the proportion of time spent on use or analysis of student assessments during professional learning and teacher preparation experiences.](chart.png)

**NOTE:** This figure depicts response data from the following survey questions. For the two bars on the left, we asked, “For the [professional development workshops or trainings/collaborative learning with other teachers, including instructional planning time] in which you have participated to support your [ELA/math] instruction this school year (2021–22), what approximate proportion of that time has been spent on [analysis or use of student assessments]?” For the two bars on the right, we asked, “What approximate proportion of your [teacher preparation program coursework, excluding your clinical teaching work/clinical teaching experience] was spent on [analysis or use of student assessments]?” The professional learning items were posed only to teachers who participated in that type of professional learning (i.e., professional development workshops and trainings or peer collaborative learning). The teacher preparation items were posed only to teachers who responded that they completed their program ten years ago or less. Columns relating to professional learning opportunities might not sum to 100 because of rounding. Columns relating to teacher preparation experiences might not sum to 100 because, when reporting on their teacher preparation experiences, teachers were also able to respond that they do not recall, which is not displayed on the graph. For this figure, we included only the responses of ELA and math teachers. From the left bar to the right bar, n = 5,656, n = 5,827, n = 1,656, n = 1,653.
Altogether, this suggests that schools and districts cannot assume that teachers’ preparation programs have adequately addressed the use of student assessment data. Even though teachers’ reports suggest that their preparation programs widely address the use and analysis of student assessments, our results suggest that ongoing support is likely necessary to help teachers understand the alignment between aspects of their instructional system or to navigate any existing misalignment.

During our interviews with teachers, about one-tenth of teachers also described how professional development or collaborative learning experiences supported them to make sense of their student data and use data to drive instruction. These experiences included meetings where teachers could get familiar with the assessments themselves and discuss student assessment results, student misconceptions, strategies to help students master standards, and opportunities

Throughout the school year, we have data meetings to kind of look at common assessments that have been created by the district . . . . [They help] us pull the data to then see where to do interventions, extensions, or where to just keep going on the same path.

— Elementary math teacher

### FIGURE 6
Percentage of Teachers Reporting Alignment Between Benchmark Assessments and Curriculum Materials by Professional Learning Experiences

<table>
<thead>
<tr>
<th>Professional Learning Experiences</th>
<th>Most or total alignment between benchmark assessments and curriculum materials</th>
<th>Partial alignment</th>
<th>Little or no alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer collaborative learning</td>
<td>More than a quarter of time spent on analysis or use of student assessments</td>
<td>68%</td>
<td>22%</td>
</tr>
<tr>
<td></td>
<td>A quarter or less time spent on analysis or use of student assessments</td>
<td>60%</td>
<td>25%</td>
</tr>
<tr>
<td>Professional development</td>
<td>More than a quarter of time spent on analysis or use of student assessments</td>
<td>69%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>A quarter or less time spent on analysis or use of student assessments</td>
<td>60%</td>
<td>25%</td>
</tr>
</tbody>
</table>

NOTE: This figure depicts response data from the following survey questions: “To what extent do the [ELA/math] benchmark assessments your students take align with the content of the curriculum materials you regularly use for your [ELA/math] instruction?” This figure includes the responses of ELA and math teachers. To determine how much time teachers spent in professional learning on analysis or use of student assessments, we asked teachers, “For the [professional development workshops or trainings/collaborative learning with other teachers, including instructional planning time] in which you have participated to support your [ELA/math] instruction this school year (2021–22), what approximate proportion of that time has been spent on [analysis or use of student assessments]?”. The professional learning items were asked only to teachers who participated in that type of professional learning (i.e., professional development workshops and trainings or peer collaborative learning). From top to bottom bar: $n = 2,700; n = 2,677; n = 2,220; n = 2,882$. Black bars represent 95-percent confidence intervals.
for remediation or extensions. Beyond their professional learning opportunities, teachers also named other factors that supported (or hindered) their effective use of student data to drive instruction (See Box 4). These data provide illustrative examples of ways that schools and districts can help teachers understand the ways that student assessments can relate to their instruction in the classroom, and these methods might be especially useful in cases where teachers may initially perceive less alignment between their benchmark assessments and curriculum materials. In other words, professional learning and resources such as those described in Box 4 can help bridge the gap between curriculum and assessment.

Although we hypothesize that professional learning opportunities—both professional development trainings and peer collaborative learning—can support teachers’ perceptions of coherence, it is possible that there are other explanations for this relationship that we observe in the data. For instance, it is possible that teachers who perceived their benchmarks as less aligned therefore invested less time in professional development or peer collaboration on analyzing and using student data. However, we are unable to determine definitively the directionality of this relationship.

Teachers’ perceptions of alignment between their benchmark assessments and curriculum materials were connected to their ability to address student needs

Our data suggest that greater perceived alignment between teachers’ benchmark assessments and the content of their curriculum materials can help them better use data to address students’ needs. Overall, among teachers whose students have taken a benchmark assessment, only 37 percent reported that when their benchmark assessments indicated that their students had unmet learning needs, they found it somewhat or very easy to identify tasks or activities within their ELA or math curriculum materials to address those needs (Figure 7). Importantly, 37 percent of teachers found it somewhat or very difficult to address the student needs identified by their benchmark assessments, and about one-quarter of teachers found it neither easy nor difficult; this is a critical finding because one goal of benchmark assessments is to help teachers adjust their instruction to meet students’ needs. Our findings thus point to an area where teachers may need additional support.

Box 4. Teachers’ Reports of Other Factors that Supported or Hindered Teachers’ Use of Benchmark Assessment Data to Meet Student Needs

Beyond professional learning, teachers named the following factors as supporting their ability to effectively use assessment data to meet student needs:

• Four teachers mentioned resources to support the link between their instructional practice and benchmark assessments, such as reports from their benchmark assessment platform to identify areas for student growth, scope and sequences or curriculum guides that outline the standards addressed in assessments, or pacing guides that explicitly note the content that should be taught in relation to the timing of the assessment.

• Two teachers mentioned data use systems. For instance, one teacher described how they could track student performance on various assessment questions and on various standards and compare their data with others in their district.

Teachers named the following factors as hindering their ability to effectively use assessment data to meet student needs:

• Three teachers mentioned unclear messaging from their leadership about how to prioritize or use different sets of assessment or assessment data.

• Three teachers talked about how they lacked professional development or other supports to use their assessment platform or to meet the standards addressed by their benchmark assessments.
This is probably the biggest, most frustrating thing of our district. They have adopted so many different ways to benchmark our kids, different formats and different resources. . . . And all of these different tests are all different standards, and they don’t align to our curriculum necessarily. . . . Looking at the data, it’s sometimes hard for us as teachers to know which data [are] important, which do I need to use to show that my children are progressing.

— Elementary ELA teacher
Engendering stronger alignment between teachers’ curriculum materials and benchmark assessments could be one way of supporting teachers to more easily use their curricula to address the student needs identified by their benchmark assessments. Teachers who perceived that their curriculum materials and benchmark assessments mostly or totally aligned were more likely than their peers who felt that there was less alignment to report that it was easy to use their curriculum materials to address student needs. Forty-seven percent of teachers who perceived alignment between their curriculum materials and benchmark assessments found it easy to address the student needs identified by their benchmark assessments, in comparison with 19 percent of teachers who perceived partial alignment and 14 percent of teachers who perceived little or no alignment. Conversely, 66 percent of teachers who felt there was little or no alignment between their benchmark assessments and curriculum materials felt it was difficult to use their curriculum materials to address students’ needs, in comparison with 26 percent of teachers who perceived that their benchmark assessments and curriculum materials were mostly or totally aligned. Although we observe these strong relationships between teachers’ perceptions of alignment and the ease with which they feel they can address student needs, we acknowledge that there are many other factors that could plausibly affect teachers’ ability to use assessment data to drive student learning, including the availability and usability of data systems, the specificity of the resulting data, and the timeliness of score reports (Herman, 2017; Marion, 2018).

Our qualitative data shed further light on this relationship between teachers’ ability to use their curricula to meet students’ learning needs and their perceptions of alignment between their curricula and benchmark assessments. When teachers perceived a gap between their curriculum materials and assessments, about one-quarter of interviewed teachers also felt the resulting assessment data were less representative of what students know and actually learned in class, and they were therefore unsure of how to use the data to drive instruction or how to use the curriculum as a vehicle to enhance student learning. As a result, teachers described the ways that they tried to engender stronger coherence between their benchmark assessments and curriculum materials (see Box 5). Conversely, there were a small number of teachers who relied on their benchmarks to provide data on students’ performance or to guide their instruction because they felt there was stronger alignment between what they were teaching and what was assessed.

Box 5. Teachers’ Reported Strategies to Promote Coherence Between Their Curriculum Materials and Benchmark Assessments

When encountering misalignment between their benchmark assessments and curriculum materials, interviewed teachers described strategies they used to navigate that perceived misalignment and instead promote coherence.

- About 20 percent of interviewed teachers described ways that they attempted to bridge what students were learning in the classroom and the content on which they were getting assessed; this was true for both perceptions of misalignment between classroom content and summative and benchmark assessments. These teachers reported either creating their own assessments that better matched their classroom lessons, therefore yielding data that teachers felt were more representative of students’ knowledge, or implementing supplemental lessons that matched the assessments so that students would be better prepared to take the assessments.
- Two teachers described the ways in which they engaged in backward planning to support the alignment between their curriculum materials and benchmark assessments. This involved thinking about the standards taught, what student mastery should look like, and how to craft lessons around those objectives.
Summary and Implications

Our report offers two insights into schools’ and teachers’ use of benchmark assessments. First, our data confirm the near-ubiquitous use of benchmark assessments in recent school years. By the 2021–2022 school year, nearly all K–12 public schools administered at least one ELA and one math benchmark assessment to at least some of their students. We observed growth across a different types of benchmark assessments, including published benchmark assessments and especially those created locally. We hypothesize that these trends are an indication of growing interest in monitoring students’ performance, especially as a response to the disruptions to data collection wrought by the COVID-19 pandemic.

Given schools’ widespread use of benchmark assessments, we also offer insight on the role that educators perceive that these assessments play in building coherent instructional systems. Benchmark assessments can be powerful tools in supporting states, districts, and schools to monitor student learning and identify learning gaps—particularly in the wake of the COVID-19 pandemic. However, it is also important that assessments are broadly aligned with other aspects of their instructional systems, conditional on their intended uses. We found that principals largely perceived strong alignment between benchmark assessments on one hand and standards and summative assessments on the other. Alignment between these factors is critical for the predictive and evaluative roles that benchmark assessments play—a use case that is common among school and district leaders. However, when we examined the linkages that we hypothesize are critical for supporting benchmark assessments’ role as instructional tools, we find that teachers largely perceive lower levels of coherence. In comparison with the roughly 80 percent to 90 percent of principals who perceived alignment between benchmark assessments and state standards and summative assessments, only two-thirds of teachers perceived alignment between their benchmark assessments and curriculum materials; furthermore, this perception of alignment between benchmark assessments and curriculum materials was strongly linked to the ease with which teachers felt they were able to address students’ unmet learning needs.

Thus, our data suggest one way to help teachers effectively use benchmark assessments to address students’ learning needs is to create stronger links between these assessments and curricula. Some teachers, such as those serving struggling learners and ELA teachers, were less likely to perceive alignment between benchmark assessments and curriculum materials—possibly because these components were misaligned or because they experience greater barriers to understanding how these elements connect.

With these two insights in mind, we offer the following recommendations.

State and local leaders should strive to create coherent instructional systems by intentionally adopting benchmark assessments aligned to other aspects of their instructional systems. These other aspects might include the standards that teachers are teaching, their required summative assessments, the curriculum materials teachers use, the timing of assessments in relation to the pacing of instruction, teachers’ professional learning schedules and opportunities, and the data systems available to teachers.

Facilitating alignment will entail trade-offs. There may be cases where education leaders see the primary purpose of benchmark assessment data as monitoring local or state initiatives or comparing student performance across schools or districts; in such cases, alignment with curricula may be less critical and alignment to state standards may be prioritized. Conversely, where schools are selecting benchmark assessments to inform teacher practice and address student misconceptions, schools or districts may instead select or create benchmark assessments that are more anchored to the content of teachers’ instruction, even though such an assessment might be less appropriate for facilitating comparisons of student proficiency.

To address these potential tensions, education leaders should consider how data from each assessment will be used from beginning to end and create a coherent set of assessments that address benchmark assessments’ varied use cases, with the understanding that some benchmark assessments may be better suited for some stakeholders and use cases.
than others. Before selecting a slate of assessments to administer, teachers and school leaders should develop a plan outlining why each assessment was chosen, how each assessment will be used, and when to administer the assessment to produce data that are well suited to its purpose. Education leaders at the state, district, and school levels might consider asking and answering such questions as: For what purpose will these assessment data be used (e.g., for teacher evaluation, guiding instruction, guiding school improvement, tracking progress on state or local initiatives)? Are the timeliness and specificity of the data well suited to their purpose? What information or facet of student learning will we understand from analyzing these assessment data? How will data from this assessment complement—or potentially contradict—data from other assessments?

A clearer vision on how benchmark assessments will be used can help education leaders and teachers better understand how they should fit within schools’ existing instructional systems and the connections among instructional system components that should be prioritized, given the assessments’ purpose and use case.

State and local leaders can also play a role in supporting the development of coherent instructional systems by holistically examining how instructional system components work together and designing or encouraging the uptake of curricula and assessments that are intentionally linked. Louisiana’s new innovative humanities assessment provides such an example and also demonstrates how a consideration of instructional coherence among multiple components of the instructional system can help mitigate the trade-offs described earlier. Beginning with the 2019–2020 school year, the Louisiana Department of Education began piloting a new state assessment which assesses students in both social studies and ELA together. The pilot initiative has resulted in the development of humanities assessments that link to two standards-aligned curricula used widely in the state, the Louisiana ELA Guidebooks and Wit and Wisdom. Because they are designed to align with teachers’ curriculum materials, they assess students on content and texts for which they’ve developed background knowledge during their regular classroom instruction (Klirs, 2022). Although the assessment is meant to replace the traditional once-a-year state summative assessment, it also has characteristics of a benchmark assessment because it is administered three times per year to allow for the generation of data that can be used to inform instruction and to provide information on student achievement against state standards (Louisiana Department of Education, undated). In describing the purpose of this new humanities assessment, the Louisiana Department of Education wrote, “This new approach is designed to make assessments more connected to the classroom for Louisiana teachers and students, while providing valid, reliable and transparent data on student achievement” (Louisiana Department of Education, 2020).

This manifestation of an assessment system that integrates connections between assessments, curricula, and standards is likely facilitated by the high uptake of standards-aligned curricula in the state—and, in particular, in Louisiana, the high statewide uptake of specific standards-aligned curricula—conditions which might not be present in all states or districts (Doan, Kaufman, et al., 2022; Louisiana Department of Education, undated). However, the rise of locally created assessments at the state, dis-
Teachers may also benefit from clearer communication behind the purpose of assessments and how they are designed to work.

district, and school levels suggest that state and local leaders have opportunities to consider how they can engender instructional coherence through the development of their own assessments while considering the trade-offs mentioned earlier.

States in particular may be well positioned to facilitate instructional coherence at local levels by incentivizing and supporting the uptake of curriculum materials and benchmark assessments that are more closely linked. Or, to the extent that state leaders find that such materials and assessments that are well suited to their contexts do not yet exist, they might also consider designing curriculum materials or benchmark assessments that are intentionally connected or encouraging partnerships between curriculum and assessment developers, as Louisiana state leaders have done. Because of the level at which state leaders sit, they are especially well positioned to consider benchmark assessments’ connection to state standards and summative assessments when designing or incentivizing uptake of certain benchmark assessments and curriculum materials.

To support the role of benchmark assessments as an instructional tool, state and local leaders should provide teachers with the professional learning and resources to facilitate connections between their curriculum materials and their benchmark assessments and to make student data more actionable. Our findings suggest that ongoing professional learning—both professional development trainings and opportunities for peer collaboration—and other resources can help teachers better connect their benchmark assessments and curriculum materials and more easily use their curriculum materials to address students’ unmet learning needs. Supports might include dedicated time to analyze and debrief on student data during professional learning opportunities, guides that help make the connection between the content and pacing of curriculum materials and assessments more explicit, resources that name the instructional practices that can address student misconceptions uncovered by assessments, or data systems that facilitate the use of student data to support instructional decisionmaking. In addition, teachers may also benefit from clearer communication behind the purpose of assessments and how they are designed to work. For instance, teachers expressed concerns or misconceptions about how computer adaptive tests evaluate student learning. Clarity around how these tests are designed to function could address teachers’ perceived misalignment between such assessments and their everyday instructional content. School- or district-level roles, such as data coaches, might also be helpful in supporting teachers to understand the connections between their benchmark assessments and other aspects of their instructional systems, such as their standards, curricula, and state summative assessments. Because teachers serving struggling learners and ELA teachers are less likely than their peers to perceive alignment, it may be possible that these activities, resources, and roles may be even more important for such teachers.

Research Limitations

This report provides a look at the prevalence of benchmark assessments from the 2018–2019 to 2021–2022 school years and examines educators’ perceptions about how benchmark assessments fit into their instructional systems. However, there are a few limitations to our research that readers should consider when interpreting results. First, our analyses are based on teachers’ and principals’ self-reports, meaning we rely on their knowledge and perceptions—a limitation present in all survey research. Because we do not provide teachers and principals with a standard definition for the constructs examined in this
asked principals about the specific benchmark assessments that their students have taken. In the 2021–2022 AIRS, we asked teachers only about two main types of benchmark assessments—a published benchmark assessment that their school or district requires them to give students and benchmark assessments that they create themselves. Because the 2021–2022 AIRS teacher survey lacked options that allowed teachers to select other kinds of benchmark assessments (e.g., school-, district-, or state-created benchmark assessments, or curriculum-embedded benchmark assessments), we lack survey data on teachers’ perceptions about those types of benchmark assessments. Moreover, it is possible that teachers may have perceived locally created or curriculum-embedded benchmark assessments as falling into the category of required, published benchmark assessments.

Finally, there are also several caveats to our analysis of the teacher interview data, which limit the generalizability of the data that we present. Because a focus on benchmark assessments was just one small part of the overall interview protocol and also because of the modest size of our sample (n = 45), our findings may not generalize to all teachers and their experiences. Instead, they are meant to be illustrative of survey findings. Furthermore, teachers self-selected to participate; therefore, we cannot generalize the findings to teachers that did not choose to participate. Although our sample is not representative, we aimed for a wide variety of perspectives and experiences. Our sample includes teachers from across the country working in a variety of grade levels and subjects, with varying years of experience and different populations of students. We focused on teachers working in schools serving a large proportion of historically marginalized students—students of color, students with individualized education plans, and English learners. Note that data on students’ eligibility for free or reduced-price lunch (a proxy for socioeconomic status) were not included in the demographic information collected on teachers, which is a limitation. These limitations are similar to the limitations of our full analysis of these teacher interviews and thus are taken from Pauketat et al., 2023.
How This Analysis Was Conducted

We leverage three sources of data—principal survey data, teacher survey data, and teacher interview data—to conduct the analyses presented in this report. In this section, we describe how we analyzed each data source to produce the results presented in this report.

Principal survey data: We used principal data from multiple years of the AIRS to examine the prevalence of benchmark assessments over time. AIRS was administered to nationally representative samples of roughly 1,500 principals in the spring of four consecutive school years (from 2018–2019 to 2021–2022) via RAND’s ASLP. Each school year, principals were asked what benchmark assessments their schools administered in that year to assess student learning. Each year, we provided principals with a list of benchmark assessments from which they could choose. Estimates for each survey were produced using cross-sectional survey weights that are designed specifically to produce nationally representative estimates in the school year that each survey was administered. A complete list of the benchmark assessments asked about on each survey are included in Table A.1 in the appendix.

We interpret principals’ responses to these survey items as speaking on behalf of their school. Therefore, we report prevalence estimates using the school as the unit of analysis (e.g., “percentage of schools who administered benchmark assessment”). However, we do note that a handful of our survey respondents (less than 1 percent) are matched to the same school, so this terminology should be interpreted with some caution.

Principals were asked separately about benchmark assessments for ELA and math. Therefore, we initially separately examined patterns in schools’ use of benchmark assessments for ELA and math. However, because results were very similar, we present data on only ELA assessments in this report for brevity, unless otherwise noted. Because results are so similar, we avoid labeling results as pertaining to “ELA benchmark assessments” and interpret the principal findings in this report as applicable to benchmark assessments in both ELA and math.

To generate prevalence estimates by school grade level, we grouped principals into three categories based on the grade levels their school served: K–8 schools (which serve students in any grades kindergarten through grade 8), high schools (which serve students in any grades 9 through 12), and span schools (which serve students in both K–8 and high school grades). We grouped middle schools in with elementary schools instead of high schools because the Every Student Succeeds Act mandates that state summative assessments be administered to students each year in grades 3–8 in both ELA and math. Because of this broader assessment framework, we might expect use of benchmark assessments in middle schools to be more similar to elementary schools than to high schools (and preliminary analyses conducted with our AIRS principal data looking specifically at middle schools confirm that this is generally true). We omit results for the small number of span schools, whose results are difficult to parse given large variance in grades served.

Teacher survey data: We used responses from 6,399 ELA and math teachers from the 2021–2022 AIRS to examine teachers’ perceptions about their benchmark assessments. We specifically focused our analysis on the items relating to teachers’ use and perceptions of their benchmark assessments, including the type of benchmark assessments that teachers administered, the extent to which the ELA or math benchmark assessments they used aligned with the curriculum materials they used regularly for their ELA or math instruction, and how difficult or easy it was for them to identify tasks or activities within their ELA or math curriculum materials to address students’ unmet learning needs identified by benchmark assessments. In addition, we analyzed teachers’ responses to questions about how much time they spent during their professional learning and teacher preparation experiences focused on use or analysis of student assessments. Additional information about each of these items is included in American Instructional Resources Surveys: 2022 Technical Documentation and Survey Results (Doan, Eagan, et al., 2022).

For both teachers and principals, we report sample-wide and subgroup-specific means and proportions of variables of interest, weighted using a set of nationally representative weights described in
To compare responses for teachers and principals in schools with different demographic profiles, we matched AIRS responses to school-level data from the Common Core of Data issued by the National Center for Education Statistics. We explored whether teachers’ and principals’ responses differed according to their demographic characteristics, their school context (e.g., school locale), or the characteristics of the students in their school. Unless otherwise noted, we reference only differences among educator subgroups that are statistically significant ($p < 0.05$).

We tested the robustness of significant differences across teacher subgroups to adjust for observable school-level characteristics (e.g., poverty level, student racial and ethnic composition, locale) and teacher-level characteristics (e.g., race, gender, and grade level). We report only differences that remained significant after controlling for teacher- and school-level characteristics. These regression analyses are useful for understanding the drivers of differences, but we do not present regression-adjusted statistics because we believe that these differences remain notable even if they could be driven by multiple underlying factors. Moreover, we did not make statistical adjustments for multiple comparisons because the intent of this report is to provide exploratory, descriptive information rather than test specific hypotheses or causal relationships.

**Teacher interview data:** In May and June 2022, a team of five qualitative researchers and analysts conducted 45 interviews with K–12 teachers. Interviewees were drawn from a nationally representative sample of ELA and math teachers who were part of the RAND American Teacher Panel in Spring 2022 and who completed a survey from a related project, the Coherent Instructional Systems project. The semistructured interview protocol focused on teachers’ perceptions about the messages and guidance that they received and the relationships between components in their instructional system, including their curriculum materials, academic standards, professional development, teacher collaboration, assessments, and teacher evaluation. Three experienced qualitative analysts then iteratively developed an initial coding scheme based on the interview protocol and, through team debriefs, established reliability, surfaced emergent codes, resolved ambiguities, and completed coding of all interviews. The findings in this report reflect analysis of only excerpts coded to assessment-related codes (see Pauketat et al., 2023, for a full reporting of results stemming from the teacher interviews). In the interviews, we asked teachers broadly about assessments (i.e., both summative and benchmark assessments), and our analysis further focused on excerpts where teachers explicitly expressed perceptions about their benchmark assessments, unless otherwise noted.
Appendix

In this appendix, Table A.1 provides basic frequency tabulations of specific benchmark assessments for interested readers.

**TABLE A.1**

Percentage of Principals Who Indicated Their Students Took Various Assessments to Assess Their Progress in ELA, by School Year

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Any locally created assessments</td>
<td>39</td>
<td>82</td>
<td>62</td>
<td>77</td>
<td>51</td>
<td>81</td>
<td>71</td>
<td>84</td>
</tr>
<tr>
<td>State-created benchmark assessments</td>
<td></td>
<td>41</td>
<td>29</td>
<td>50</td>
<td></td>
<td>49</td>
<td>32</td>
<td>49</td>
</tr>
<tr>
<td>District-created benchmark assessments</td>
<td>30</td>
<td>56</td>
<td>42</td>
<td>50</td>
<td>29</td>
<td>47</td>
<td>34</td>
<td>46</td>
</tr>
<tr>
<td>School-created benchmark assessments</td>
<td>21</td>
<td>45</td>
<td>25</td>
<td>35</td>
<td>36</td>
<td>53</td>
<td>41</td>
<td>49</td>
</tr>
<tr>
<td>Any published assessments</td>
<td>80</td>
<td>84</td>
<td>86</td>
<td>87</td>
<td>52</td>
<td>62</td>
<td>56</td>
<td>62</td>
</tr>
<tr>
<td>Achievement Network (ANet) Interim Assessments</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Achievement Series (Scantron)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ACT Aspire (ACT, Inc.)</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>24</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Acuity (CTB/McGraw Hill)</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Adaptive Assessments (Edmentum)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ascensus (Scantron)</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>aimswebPlus (Pearson)</td>
<td></td>
<td>10</td>
<td>9</td>
<td>9</td>
<td></td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Collaborative Assessment Solutions for Educators (CASE)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Compass Learning (EdGenuity)</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Continuum (Houghton Mifflin Harcourt)</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CTB Assessments (McGraw Hill)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Developing Skills Checklist (Buros Center for Testing)</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Developmental Reading Assessment (DRA) (Pearson)</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Dynamic Indicators of Basic Early Literacy Skills (DIBELS)</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Discovery Education (Discovery Education)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>eMPower (Measured Progress)</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Exact Path (Edmentum)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>FAST (FastBridge Learning)</td>
<td>—</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>—</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Fountas &amp; Pinnell Benchmark Assessments</td>
<td>19</td>
<td>19</td>
<td>17</td>
<td>15</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Gates-MacGinitie Reading Tests (Houghton Mifflin Harcourt)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Any i-Ready</td>
<td>18</td>
<td>24</td>
<td>26</td>
<td>29</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>11</td>
</tr>
</tbody>
</table>
Table A.1—Continued

<table>
<thead>
<tr>
<th></th>
<th>K–8 Schools</th>
<th>High Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>i-Ready Assessments (Curriculum Associates)</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>i-Ready Diagnostic (Curriculum Associates)</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>i-Ready Standards Mastery (Curriculum Associates)</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>iSTEEP (ISTEEP)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MAP or Measures of Academic Progress (NWEA)</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>mCLASS (Amplify)</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Performance Series Assessments (Scantron)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Pivot-INSPECT Benchmark Assessments</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Previ Learn (CenterPoint)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Riverside Interim Assessments (Houghton Mifflin Harcourt)</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Any Smarter Balanced</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Smarter Balanced (SBAC) interim assessment bundles</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Smarter Balanced (SBAC) interim comprehensive assessments</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Any Star</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Star Assessments (Renaissance)</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Star Reading/Star Math (Renaissance Learning)</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>Study Island (Edmentum)</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Pre-made interim assessment from an item bank</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Free online assessments</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>We use a benchmark assessment, but I don’t know the name of it</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>N/A—My students do not take a benchmark assessment for this subject</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

NOTE: — indicates that data are not available because our survey did not ask about that specific benchmark assessment in that school year. This table uses responses to the following survey item administered to school principals: “Which benchmark assessments do your students take over the course of this school year [school year] to assess their progress in ELA and mathematics?” Respondents were instructed to “select all that apply.”
Notes

1 Perie, Marion, and Gong, 2009, which is a seminal piece in the literature on interim assessments, notes that assessments referred to as benchmark, diagnostic, or predictive are included in their definition for interim assessments. Because we asked teachers and principals about their use of benchmark assessments, we use the term “benchmark” throughout this report and use it interchangeably with “interim.”

2 Our survey asked principals separately about “i-Ready Assessment,” “i-Ready Diagnostic,” and “i-Ready Standards Mastery.” We use data from these three subitems to determine which schools used “any i-Ready” assessment.

3 Our survey asked principals separately about “Star Assessments (Renaissance)” and “Star Reading/Star Math (Renaissance Learning).” We use data from these two subitems to determine which schools used “any Star” assessment.

4 Respondents’ schools also may have administered the same assessment multiple times in the same school year (e.g., administering MAP Growth in the fall, winter, and spring of the school year). However, in our data, this is considered administration of a single benchmark assessment (as opposed to three benchmark assessments).

5 Trends were constructed using cross-sectional estimates produced independently using the AIRS American School Leader Panel (ASLP) survey sample in that school year. AIRS uses a probability-based sampling and weighting scheme to ensure that survey estimates are representative of the national population of public school principals in that survey year. (For more information see Doan, Eagan, et al., 2022.) While we note that survey completion rates varied to some degree across survey administrations (the 2018–2019 survey had a 32.5 percent completion rate, the 2019–2020 survey had a 28.8 percent completion rate, the 2020–2021 survey had a 35.7 percent completion rate, and the 2021–2022 survey had a 39.7 percent completion rate), we are not very concerned about bias in sample composition between survey years driving these trends, given the AIRS sampling and weighting procedures.

6 Because we administered the 2019–2020 AIRS about two months after school closures related to the COVID-19 pandemic began in March 2020, our data include assessments administered after school closures began. However, we acknowledge that some schools likely had already administered benchmark assessments before their schools closed.

7 Principals could indicate that their school administered multiple benchmark assessments but were not asked about alignment separately for each specific benchmark assessment their school administered.

8 We defined teachers as regularly using curriculum materials if they used them once a week or more on average, while we defined teachers as intensively using curriculum materials if they used them once a week or more on average and used them for 50 percent or more of their instructional time for a typical class of students over the course of a week.
About This Report

In this report, drawing on the American Instructional Resources Survey (AIRS), the authors examine the prevalence of different types of benchmark assessments from the 2018–2019 to 2021–2022 school years using principals’ reports. Using the 2021–2022 AIRS and interviews with 45 English language arts and math teachers, we also explore educators’ perceptions about the extent to which benchmark assessments align with other components of their schools’ instructional systems. RAND’s American Educator Panels provide access to nationally representative samples of K–12 teachers, principals, and district leaders across the country. The panels are a proud member of the American Association for Public Opinion Research’s Transparency Initiative. If you are interested in using AEP data for your own surveys or analysis or in reading publications using American Educator Panel data, please email aep@rand.org or visit www.rand.org/aep.

RAND Education and Labor

This study was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through post-secondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking. This report is based on research funded by the Bill & Melinda Gates Foundation, the Charles and Lynn Schusterman Family Foundation, and the Walton Family Foundation. The findings and conclusions presented are those of the authors and do not necessarily reflect positions or policies of the foundations that supported this research. More information about RAND can be found at www.rand.org. Questions about this report should be directed to awoo@rand.org, and questions about RAND Education and Labor should be directed to educationandlabor@rand.org.

Acknowledgments

We are extremely grateful to the U.S. public school teachers and leaders who agreed to participate in the panels. Their time and willingness to share their experiences are invaluable for this effort and helping us understand more about how to better support their hard work in schools. We thank Alvin Nugroho for serving as the survey manager and Casey Hunter for serving as the data manager for this survey, and Tim Colvin for programming the survey. Thanks to Claude Messan Setodji for producing the sampling and weighting for these analyses. We also greatly appreciate the administrative support provided by Tina Petrossian. We also thank our reviewers, Jon Schweig and Megan Kuhfeld, for helpful feedback that improved this report. Lastly, we thank Anna Bloom for her editorial expertise and Monette Velasco for overseeing the publication process for this report.