Research Report

Identifying Teachers' Needs for Results From Interim Unit Assessments

ETS RR–21-08

Priya Kannan
Andrew D. Bryant
Shiyi Shao
E. Caroline Wylie

December 2021

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Identifying Teachers' Needs for Results From Interim Unit Assessments

Priya Kannan, Andrew D. Bryant, Shiyi Shao, & E. Caroline Wylie

Educational Testing Service, Princeton, NJ

Interim assessments have been defined variously in different contexts and can be used for predictive purposes or instructional purposes. In this paper, we present results from a study where we evaluated reporting needs for interim assessments designed for instructional purposes and intended to be used at the end of defined curriculum units. Results from such unit assessments should help teachers determine gaps in student understanding and inform ongoing instructional decision-making. Our goal was to determine if learning progressions (LPs) could serve as the cognitive lens through which teachers can evaluate how their students' understanding of key constructs improves through periodic unit assessments. Therefore, we used the LP framework in mathematics and the key practices (KP) framework for English language arts (ELA) to design preliminary teacher report mock-ups for these unit assessments. Within a utilization-oriented evaluation framework, we conducted six needs-assessment focus groups with elementary and middle school mathematics ($n = 12$) and ELA ($n = 11$) teachers to specifically evaluate the extent to which they find results presented within the LP and KP frameworks understandable and useful for their instructional practice. Results from the focus groups show teachers' overall needs for types of information sought from unit assessment reports, the extent to which teachers are familiar with the LP and KP frameworks, their interpretations (including confusions) of the information presented in the preliminary mock-ups, and their additional needs for reports from unit assessments to be instructionally useful.

Keywords Interim assessments; teacher reports; student reports; family reports; learning progression; key practice; mathematics; English language arts; elementary school students; middle school students

doi:10.1002/ets2.12320

Teachers constitute a critical stakeholder group, particularly in K – 12. Throughout the school year, teachers assess students through an array of formative, interim, and summative, or end-of-year, assessment and desire actionable feedback about their students' performance across the range of assessment formats. However, different types of assessments are used to gather information about student learning at different grain sizes and for different purposes (Wylie & Lyon, 2017). Consequently, reports of results also tend to include different information at varying grain sizes. Even though teacher reports should always be intended to help identify gaps in students' understanding and address those gaps, the assessment context often dictates teachers' specific needs for results in the score reports. For example, while teachers seek a snapshot of a student's current level of understanding and their readiness for the next grade from end-of-year accountability assessments, they require more detailed on-going feedback about the gaps between their students' current level of understanding and specific learning goals from formative assessment (Bennett, 2011; Black & Wiliam, 1998).

Interim assessments fall somewhere between formative and summative assessment and have been defined variously. Interim assessments can either be intended for predictive purpose (to predict student's performance on the annual end-of-year accountability assessment) or be intended to identify and provide appropriate instructional interventions (Perie et al., 2007). The more instructionally focused interim assessments are intended to be used at the end of defined curriculum units to identify students' conceptual understanding of key constructs in the domain. Such assessments have also been variously called "unit assessment" or "common assessments." In recent years, the use of such unit assessments has considerably increased within classrooms (Goertz et al., 2009) because they are expected to determine gaps in student understanding and meaningfully inform instructional practice for teachers.

However, these anticipated benefits of unit assessments inherently depend on the data/results reported from these assessments. In order for unit assessments to lead to greater differentiation in instruction and to lead to improvements

Corresponding author: P. Kannan, E-mail: pkannan@ets.org

in student achievement, the results reported for these assessments should provide teachers with appropriate and actionable feedback. Particularly, the reports should help teachers with student grouping and provide suggested next steps and resources that are targeted to specific students' levels of understanding (Shute, 2008; Zapata-Rivera et al., 2007). In this paper, we will present the results from a study where we evaluated teachers' needs for results from interim assessments that are designed to be used at the end of defined curriculum units and aligned to learning progressions (LPs) in mathematics and key practices (KPs) in English language arts (ELA; see Bennett, 2010).

Learning Progressions and Key Practices Frameworks

LPs are defined as "successively more sophisticated ways of reasoning within a content domain" (Smith et al., 2006, p. 1). LPs are empirically grounded and testable hypotheses about how students begin to understand core concepts and how this understanding becomes more sophisticated over time with experience and instruction. As defined, LPs do not have an inherent grade-level expectation; rather, learning within each construct is conceived of as progressively sophisticated understandings of that construct. LPs describe how knowledge and skills typically change with respect to central ideas of a domain as students' competency improves.

In order for teachers to assess each student's current level of understanding and identify gaps or partial understandings (Smith III et al., 1994), teachers would need to understand the continuum underlying learning within any particular knowledge domain (Heritage, 2008). LPs are clearly articulated to reflect the progression of understanding within a domain and can provide the big picture of what is known and what is to be learned; thereby, they inherently support teachers' instructional planning (Heritage, 2008). The levels within an LP are derived based on empirical research and represent the levels of conceptual understanding and development. Earlier levels may reflect mistaken or incomplete understandings of the concepts that have to be revised or abandoned to achieve a more complete understanding reflected by higher levels on the progressions (Corcoran et al., 2009).

One example of an LP for proportional reasoning is presented in Figure 1. It can be seen from Figure 1 that the LP for proportional reasoning describes the progression of learning from a naive or simplistic understanding of that concept (that may include misconceptions) to a robust and expert level of understanding. When teachers understand the underlying developmental continuum of partial understandings, it may be easier to identify targeted intervention for individual students and/or small groups to help guide them toward higher levels of understanding in the domain. Therefore, tasks

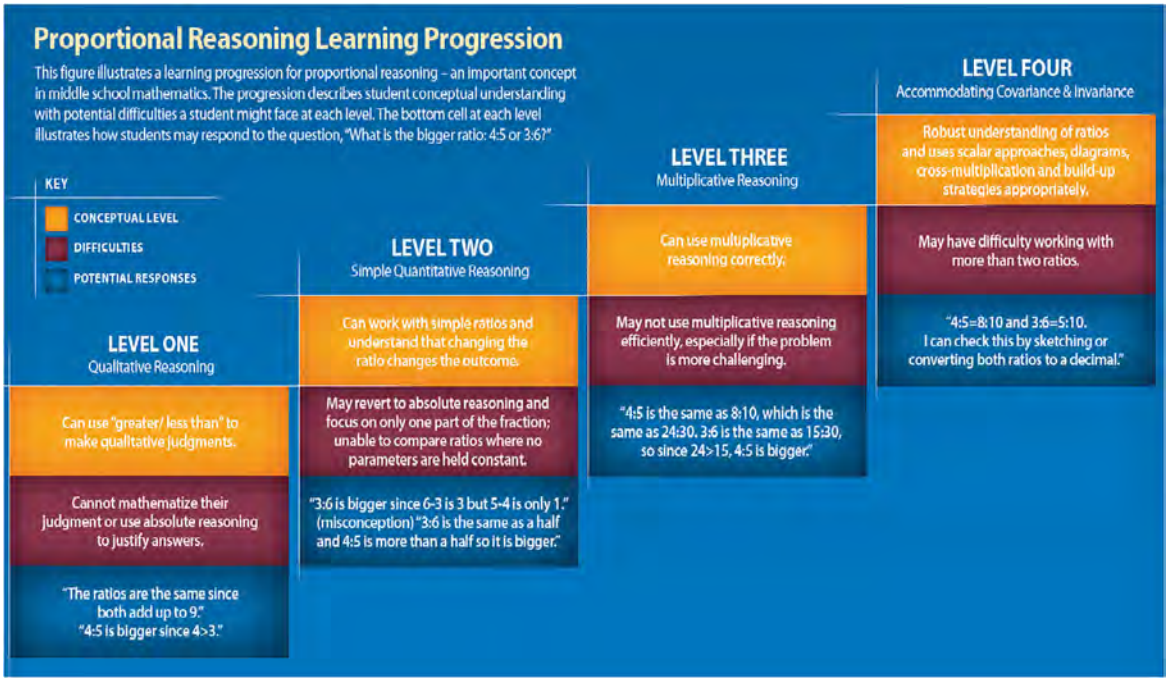

Figure 1 Illustration of a learning progression for proportional reasoning in mathematics.

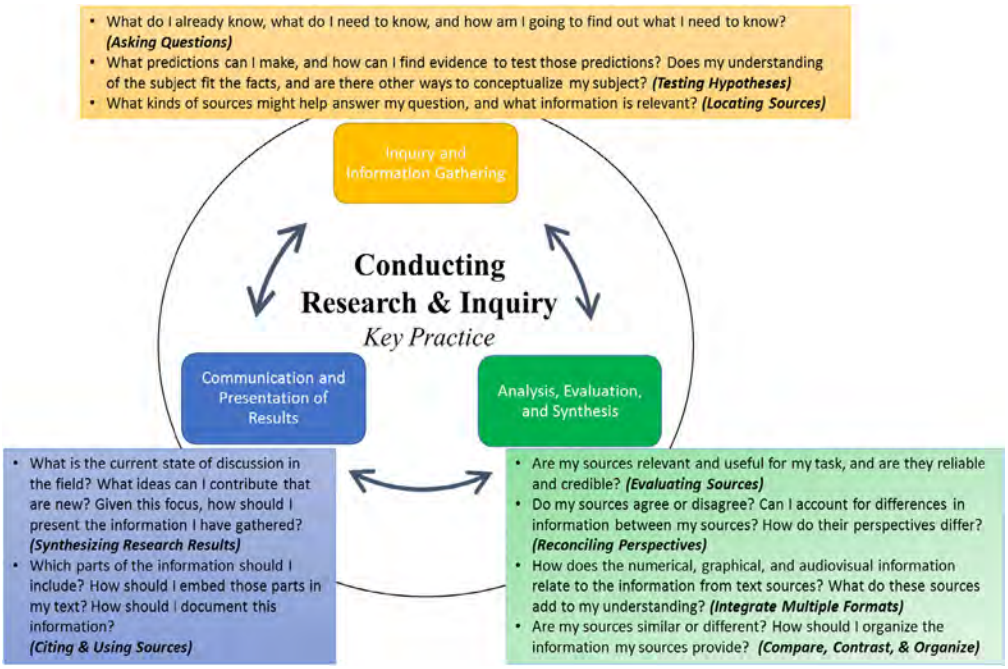

Figure 2 Illustration of a key practice for conducting research and inquiry in English language arts.

may be structured to target various levels of the progression so teachers can better identify what students know and are struggling with and provide suggestions and information to move that learning forward.

Within ELA, KPs have been used as an organizing structure for a set of closely related tasks. These KPs are presented as a sequence of phases that students can move through in order to accomplish the primary goal (e.g., discussing and debating ideas, conducting research and inquiry) described by the KP (see Figure 2 for an example of the phases contained within the key practice of Conducting Research and Inquiry). The KP framework draws heavily from literature on cognition and sociocultural theories of reading and writing and focuses on the ways in which specific skills work in concert to support the rich range of literacy practices (Deane et al., 2015). The KP framework focuses on defining categories of activities and skills and mapping these onto LPs. The KP framework provides a comprehensive description of the full developmental continuum of ELA knowledge and skills from preschool through high school (Deane et al., 2015).

Interim Unit Assessments Used

The unit assessments used in this study are being designed as modular, scenario-based assessments to measure student performance with respect to state standards and aligned to LPs in mathematics and KPs in ELA for defined curriculum units. These assessments are being developed with an intention to help teachers understand their students' current conceptual understanding within a domain, identify gaps before and after instruction, and decide on the appropriate pedagogical actions including individualized and grouped instructional next steps (Heritage, 2008). Therefore, it is conceivable that providing feedback within a framework of LPs and KPs could serve as the "cognitive lens" (Heritage, 2008) through which teachers can evaluate students' current levels of conceptual, procedural, and strategic understanding of key constructs. Such feedback should be designed to help teachers not only identify students' current level of understanding, but also specifically identify areas of underlying misconceptions or partial understandings within the domain and provide insights to teachers about instructional next steps (Shepard, 2018).

Each of these unit assessments are being designed to discriminate student performance on two or more levels of the LP or KP relevant to the domain and will take a single class period to administer. They are intended to provide teachers with the flexibility of administering a unit assessment targeted to their instructional goals (i.e., at the point in time when instruction of the relevant content is addressed in the classroom) and targeted to each student's current level of understanding. For example, as a teacher is introducing concepts around linear functions for the first time in Grade 6, they may choose to administer a unit assessment that is designed for Levels 1 and 2 of the linear functions LP, and as instruction

progresses (both within Grade 6 and in later grades), teachers may choose to administer unit assessments aligned to higher levels of the linear functions LP. Teachers may also choose to administer unit assessments before beginning a unit (i.e., preunit assessment), mid-way during the unit (i.e., mid-unit assessment), or after instruction on a curriculum unit is completed (i.e., end-of-unit assessment). At each of these time points, the teacher may choose different levels of the assessment for different students in the class. For example, if the teacher has a reason to believe that some students have advanced further than others in their understanding of linear functions by the end of the curriculum unit, the teacher can choose to administer a Levels 1 – 2 assessment for some students and a Levels 3 – 4 assessment for other students.

Score Reports for Teachers

Most of the research on score reporting, including research around how teachers make sense of data (Mandinach & Gummer, 2016; Marshall & Drummond, 2006; Zapata-Rivera, 2011b; Zapata-Rivera et al., 2010; Zapata-Rivera et al., 2016; Zwick et al., 2008) has focused on summative assessment results. As pointed out earlier, the context and purpose of the assessment (e.g., summative versus formative) tend to dictate the specific types of information that would interest and be meaningful for teachers. Teachers require high-quality and timely feedback particularly from interim and formative assessments when results can still be used to inform instructional and curriculum decisions (Underwood et al., 2007). This requirement has been referred to as "who needs to be taught what next" (Brown et al., 2019, p. 109).

This particular teacher need especially rings true in the formative and unit assessment contexts (Black & Wiliam, 1998; Kulik & Kulik, 1988; Nicol & MacFarlane-Dick, 2006; Sadler, 1989) where teachers are occupied with ongoing instructional planning. The timeliness of feedback is very important, and feedback provided in real time has been claimed to produce immediate, noticeable gains (Kulik & Kulik, 1988). In particular, students taking on novel or difficult tasks appear to benefit from immediate feedback in overcoming misunderstandings (Shute, 2008). Therefore, teacher score reports from unit assessments should be designed to specifically support gaps in student's conceptual understanding, be able to inform student groupings, and provide suggested next steps to tailor instruction (Shute, 2008; Zapata-Rivera et al., 2007). Ideally, reports designed for such unit assessments should provide teachers with a tool kit for instructional planning with results presented to support appropriate instructional decision-making, interventions, and actions.

However, from several decades of research on teacher comprehension of information presented in summative score reports, we know that teachers often struggle to parse out some of the technical information presented in score reports (Hambleton & Slater, 1997; Impara et al., 1991; Underwood et al., 2007; Zapata-Rivera et al., 2012). Moreover, teachers have reported that they are often inundated with large volumes of raw student data, which they find overwhelming to sift through. This phenomenon is referred to as "data rich – information poor," or DRIP, which was first proposed in the field of healthcare (Goodwin, 1996) and later extended to refer to the overwhelming amounts of data available to educators (Charman, 2009) in today's context of ever-increasing federally mandated standardized assessments. Teachers' experience of being overwhelmed is evidenced in a study by Zapata-Rivera et al. (2012), where they found that teachers preferred short, easy-to-read pieces of information in evaluating their students' performance on summative assessments. Therefore, teachers' reports designed for unit assessments should not be inundated with too many pieces of data; rather, the focus should be on easily summarized and instructionally meaningful pieces of information.

Iterative Multistep Score Report Development Framework

The importance of score reports as the primary vehicle for communicating both assessment purpose and results to stakeholders has been increasingly acknowledged. As Zapata-Rivera and Katz (2014) put it, score reports are truly where "the rubber hits the road" (p. 442). Over the last couple of decades, research on score reporting has highlighted the importance of considering an assessments' purpose and claims as score reports are designed. Consequently, some iterative multistep frameworks (Hambleton & Zenisky, 2013; Tannenbaum et al., 2016; Zapata-Rivera, 2011a) have been proffered for the design and development of score reports. With an overall goal of communicating assessment results in ways that allow stakeholders to appropriately interpret and use assessment results to inform their decision-making, these frameworks emphasize an agile and iterative model to develop audience-specific score reports.

Figure 3 offers a visual illustration of the recommended multistep framework for report design and development. As can be seen from Figure 3, the recommended process is fairly iterative starting from a consideration of the test's purpose and claims to develop some preliminary snapshots or mock-ups. These preliminary mock-ups are designed by preempting

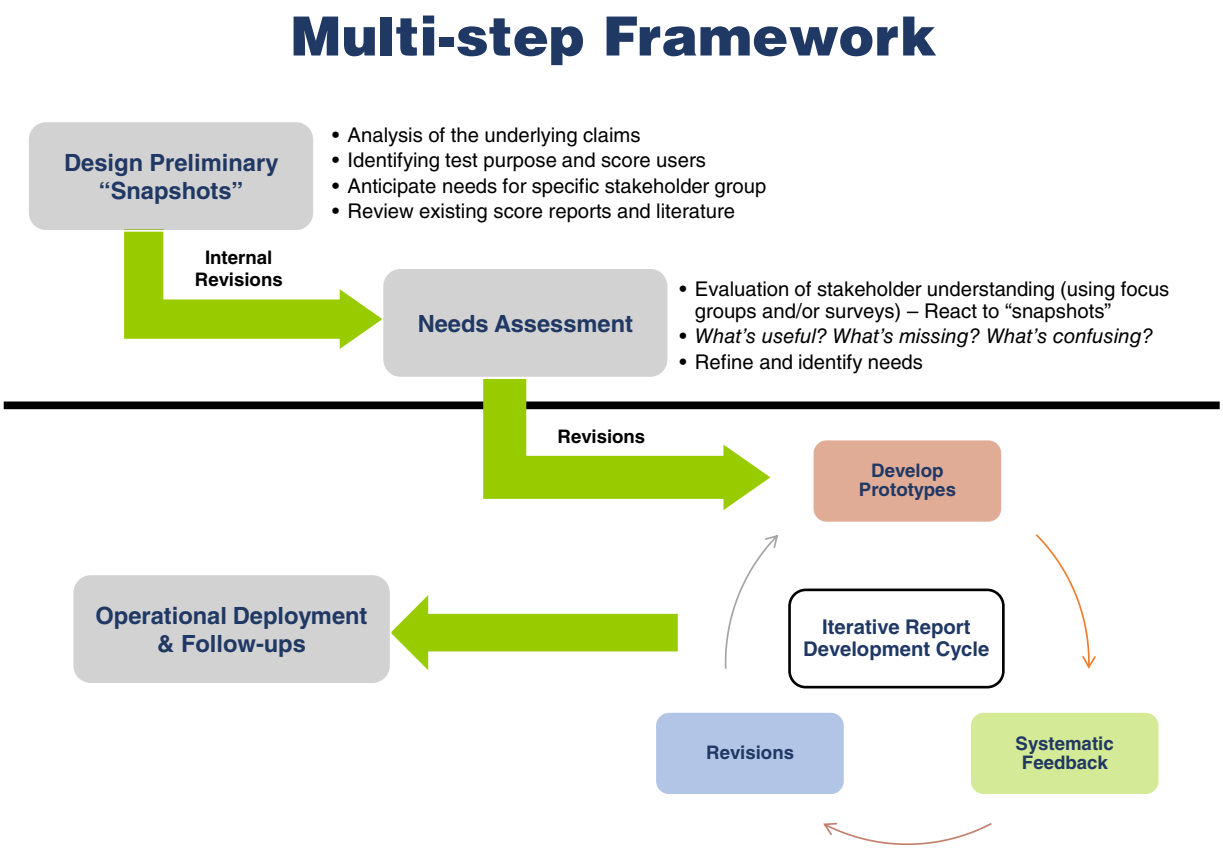

Figure 3 Recommended multistep framework for designing audience-specific score reports.

stakeholder needs prior to eliciting audience-specific needs (Zapata-Rivera & Katz, 2014) by conducting a needs assessment study. Once stakeholder-specific needs are identified through organized focus group activities during the needs assessment, the multistep framework continues through ongoing and iterative score report development and evaluation cycles, where functional prototypes are iteratively evaluated for comprehension and usability (e.g., Kannan et al., 2018) and revised before operational deployment of the score reports.

In this study, we focused on the first phase of this multistep framework (see Figure 3) and conducted a series of needs-assessment focus groups with mathematics and ELA teachers from around the country. We designed a series of preliminary reports for the interim unit assessments, keeping in mind underlying principles of visual communication and possible stakeholder knowledge and attitudes (Hattie, 2009; Hegarty, 2019; Tufte, 2001). Overall, we were interested in eliciting teachers' needs for feedback from unit assessments and worked within a utilization-oriented evaluation framework (Greene, 1988) where stakeholders' participation is elicited at every step of the evaluation process. Moreover, we also wanted to evaluate the extent to which teachers find the preliminary snapshots understandable and useful for their instructional practice—we specifically focused on what they liked and disliked within these snapshots and what they found confusing. The methods used and results from these needs-assessment focus groups are presented in the rest of this paper.

Method

Research Questions

We designed preliminary mock-ups of report screens for unit assessments in ELA and mathematics (see Appendix A). The broad purpose of this study was to evaluate those preliminary mock-ups with the intended stakeholder group (i.e., teachers) and elicit any additional needs for feedback that teachers may have in this context. We conducted six needs-assessment focus group meetings with elementary and middle school ELA and mathematics teachers from across the United States.

We addressed the following broad research questions in this study: (a) What types of unit assessments do these teachers currently use? What types of reports, if any, do they receive for these assessments? (b) What are the top needs for results from unit assessments for elementary and middle school ELA and mathematics teachers? (c) To what extent are the results presented in an LP/KP framework comprehensible and useful to teachers? (d) To what extent do the preliminary mock-ups address teachers' most important needs for results from unit assessments? What is missing and what could be improved?

In addition to the above questions, we were also interested in finding out more about the unit assessment use case. Specifically, we were interested to find out if teachers (both mathematics and ELA) used unit assessments predominantly at the end of curriculum units or if unit assessments were commonly used both at the beginning of new curriculum units (i.e., as preunit assessments) and at the end of those units. Finally, we wanted to understand the extent to which teachers would find targeted links to curated and open educational resources (OERs) helpful. We specifically wanted to find out the extent to which they were familiar with some commonly used curated and OERs (i.e., Khan Academy, OpenEd, OER Commons, Curriki, and Education.com), if they would like recommendations for any of these OERs included in the unit assessment reports, and if they would recommend other OERs that they had used in their context.

Materials

Preliminary Mock-Ups

In accordance with the recommendations in the literature (Shute, 2008; Tannenbaum et al., 2016; Zapata-Rivera, 2011a), we designed preliminary mock-ups taking into consideration the unit assessment's informational purpose, recommended principles of visual communication (Hattie, 2009; Hegarty, 2019; Tufte, 2001), and possible stakeholder knowledge and attitudes (Zapata-Rivera & Katz, 2014). These preliminary mock-ups (see Appendix A) were designed to provide actionable feedback to teachers (e.g., help with student groupings [see Figures A2 and A7] and provide instructionally useful suggestions for next steps and resources [see Figures A3 and A8]). We specifically used the LP framework (see Figures A3 and A8) to organize unit assessment results for mathematics and the KP framework to organize results for ELA. We hoped that the LPs and KPs would serve as a cognitive lens (Heritage, 2008) through which teachers can evaluate how student's conceptual, procedural, and strategic understanding of key constructs improves through periodic unit assessments. Therefore, we hoped to provide teachers with evidence about students' profiles of partial understandings within each LP (see Figures A3 and A8) and evidence of students' progress in understanding from the beginning of the unit to the end of the unit presented in LP levels (see Figures A4 and A9). These mock-ups were also informed by preliminary market research where teachers and other stakeholders reacted positively to the concept of narrative feedback (see Figures A2 and A7) about students' current level of understanding. Once these mock-ups were designed as a logical next step within the multistep framework (Figure 3), we then evaluated these preliminary mock-ups with the intended stakeholders (i.e., mathematics and ELA teachers) through a series of six needs-assessment focus groups (virtual and face to face).

Instruments

To inform the discussions within the focus groups, a premeeting task was used. We specifically wanted to understand the types of feedback teachers received (or created themselves) for unit assessments currently used within their classrooms. This premeeting task was completed online prior to the start of the study through SurveyGizmo, an online survey tool. A link to the survey was emailed to the participants about a week before each scheduled meeting. A semi structured format, with whole-group discussions interspersed with individual evaluative activities, was used for the focus groups. Study procedures are detailed in a later section.

Participants

We conducted six focus groups (three mathematics and three ELA) with elementary and middle school mathematics ($n = 12$) and ELA ($n = 11$) teachers to evaluate the extent to which they found results presented within these preliminary mock-ups (i.e., within an LP and KP framework) understandable and useful for their instructional practice. In addition, we elicited teacher-specific needs for results from unit assessments. Participant characteristics are presented in Table 1. The first focus group for mathematics was conducted as a face-to-face focus group in Princeton, New Jersey, with local

Table 1 Participant Characteristics for Mathematics and English Language Arts (ELA)

Characteristics	Mathematics ($n = 12$)	ELA ($n = 11$)
Gender		
Female	6 (50%)	7 (64%)
Male	6 (50%)	4 (36%)
Ethnicity		
Prefer not to answer	1 (8%)	
Black or African American	3 (25%)	1 (9%)
Hispanic/Latino		1 (9%)
American Indian	1 (8%)	1 (9%)
White, non-Hispanic	7 (58%)	8 (73%)
Years of experience		
Less than 7 years	2 (16%)	4 (36%)
8 – 15 years	3 (25%)	5 (45%)
16 years or more	7 (58%)	2 (18%)
Grade level teaching		
Kindergarten – fourth grade	4 (33%)	5 (45%)
Fifth and sixth grade	3 (25%)	3 (27%)
Seventh and eighth grade	3 (25%)	2 (18%)
Ninth and 10th grade	2 (16%)	
Reading specialist		1 (9%)
Subjects taught		
Mathematics	8 (67%)	—
ELA	—	6 (55%)
Mathematics and ELA	3 (25%)	5 (45%)
Region		
South	5 (42%)	7 (64%)
Northeast	6 (50%)	1 (9%)
Midwest	1 (8%)	—
West	—	3 (27%)

(New Jersey and Pennsylvania) mathematics teachers. The remaining five focus groups (two more for mathematics and three for ELA) were all conducted online as virtual focus groups through Zoom video conferencing and included teachers from various regions of the country. All focus groups were conducted between June 1 and June 28, 2018.

Procedures

Participants in all six focus groups first completed a premeeting assignment where they were asked about the types of unit/common assessments they used and the types of reports they created or received for those assessments. Teachers' responses to the premeeting task were compiled prior to the focus group and used to start the conversations during the first whole-group discussion. In addition, participants in the five virtual focus groups were asked about their familiarity with five popular OERs during the premeeting assignment. This question was added after the first face-to-face focus group because we did not receive clear response from participating teachers about their use of and familiarity with online resources during the first meeting.

The virtual focus groups were conducted using Zoom meeting software, which offered us the flexibility of scheduling participants who could participate in a single focus group without the need for travel. Combining the benefits of an online meeting scheduling software with online survey participation tools such as SurveyGizmo enabled us to provide participants with a close to real-life experience of focus group participation. The survey tools were embedded within various sections of the meeting to procure real time feedback from participants as we would in a face-to-face meeting. In addition, a heat-mapping tool was used to help participants mark up the wireframe screens in a manner similar to a face-to-face experience of marking up a wireframe. On average, we scheduled only four or five participants in each virtual focus group so that these sessions could be completed within a reasonable amount of time (about 3 hr) and so that all participants had an opportunity to contribute to the group discussions.

During all focus groups, participants were presented with a description of the nature and claims of the unit assessments being developed. Subsequently, participants engaged in a series of whole-group discussions and independent evaluative

activities that were intended to identify their needs for results from these unit assessments and the extent to which they were able to understand the information presented within the preliminary mock-ups.

The first whole-group discussion was intended as a follow-up to the premeeting assignment that was intended to evaluate their current practice. We also wanted to engage the participants in an open discussion of their needs for results in this context prior to showing them the preliminary mock-ups and eliciting their reactions to those mock-ups. Participants were also asked to discuss potential use cases (i.e., preunit versus end-of-unit use case) for unit assessments during this first whole-group discussion. Post discussion, we asked participants to reflect upon the preceding discussion and independently identify their top five needs for results in this context.

Once all participants had identified their top five needs, we presented the preliminary mock-ups designed for this study (see Appendix A). Participants were asked to work independently and mark up each report screen indicating what they liked, what they disliked, and what they found confusing. Next, they were asked to evaluate all report screens as a set and consider if their individual top needs were addressed within these screens, and if not, they were asked to indicate what types of needs were not addressed or missing from the current set of screens. After these activities were completed independently, the focus group meetings were wrapped up with another whole group discussion where the top needs across all participants were reviewed and discussed with particular focus on anything that was prominently missing from the current snapshots.

Finally, we engaged the panelists in a discussion about open and curated educational resources available to teachers. We asked participants to indicate not only their familiarity with the five resources listed in the premeeting materials, but also the extent to which these resources were particularly useful in their context. In addition, they were asked to identify any additional OERs (beyond the five identified in the premeeting assignment) that they may have used and to describe what they found useful about these OERs.

Results

In this section, results from the six focus groups are consolidated and presented to answer the broad research questions evaluated in this study; within each subsection, results are presented separately for the mathematics and ELA focus groups.

Current Practices in Unit Assessments

To understand teachers' current practices in creating, administering, scoring, and receiving/creating reports for unit assessments, we asked teachers to reflect on a few questions during the premeeting work. We reviewed their responses and discussed them further during the focus group meetings. This section summarizes the types of unit assessments teachers currently used and the types of feedback they received (or created) for these assessments.

Assessments Used

Both ELA and mathematics teachers said that, for the most part, they tended to develop, score, and compile results for their own unit assessments. Occasionally, they complement these teacher-developed assessments with some externally developed (or off-the-shelf) assessments for which they obtained structured reports. We wanted to understand the extent to which either of these practices were prevalent within the current sample of teachers.

Mathematics

All of the mathematics teachers ($n = 12$) reported that they used some type of teacher-developed assessments for their unit assessments. They indicated that they developed these questions based on curriculum standards (either the Common Core or state-specific standards) followed within their state. Most teachers mentioned that they developed these assessments either in collaboration with their grade-level teams or by themselves using a test-generator software that often came with their curriculum or textbook. Sometimes, they also reused unit assessment items from previous years or released or practice items from standardized summative assessments. Almost all teachers indicated that even though they liked administering unit assessments to their students to assess what they had taught and to tailor the learning for their students, they disliked the amount of time it took to create their own unit assessments and found it cumbersome to create and score

assessments for each unit. Eight of the 12 mathematics teachers also reported using one or more externally developed unit assessments. Specifically, they mentioned using the following off-the-shelf assessments: Math Expressions, EnVision Math series by Pearson, Eureka Math, Mastery Connect, Everyday Math, District Benchmark Assessments, iReady assessments, Street Math, and McGraw Hill Connect ED (Envision Math, Eureka Math, Everyday Math, and iReady were each named by two or more teachers). Teachers indicated that they liked these assessments because they were premade, involved considerably less work for the teachers, were directly aligned to standards, and provided consistency across schools within a district.

English Language Arts

For ELA, eight of the 11 teachers indicated that they developed unit assessments either in isolation or in collaboration with their grade-level teams. These teacher-developed assessments were intended to assess either their students' reading comprehension skills based on passages or their mastery of vocabulary or literacy skills after the conclusion of related units. Like the mathematics teachers, the ELA teachers also appreciated the ability to control the content of these teacher-developed assessments to their specific instructional goals, but again, they disliked the amount of time it took to develop and score these assessments. Eight of the 11 ELA teachers also reported using one or more externally developed (i.e., off-the-shelf) unit assessments. Specifically, they mentioned using the following off-the-shelf assessments: CTP4 by ERB, Membean (for Vocabulary), NoRedink.com, Readers and Writers, iReady assessments, Leap 360 assessments, state-released practice tests, NWEA STAR, EAGLE, Study Island, Pearson Reading Street, Pearson DRA2, and DORA. Overall, teachers said that they liked the convenience of downloading and using off-the-shelf unit assessments to easily determine if their students are mastering content. Positive features include the fact that these assessments were computerized and, in some cases, that the questions were in the same format as the state assessment. Teachers also liked that some of these off-the-shelf unit assessments gave them the ability to monitor progress through the school year for individual students, classrooms, and grades across their school and district.

Reports Received or Created

We asked teachers what types of results or reports they currently received for the off-the-shelf unit assessments they used. We also asked teachers if they created any summaries or reports for the teacher-developed assessments.

Mathematics

Teachers who developed their own assessments said that they mostly recorded student scores on these assessments in a gradebook; these scores were then reflected in student report cards. In addition, those who used software that came with their curriculum or textbook to generate items indicated that they could use the same program to analyze data and compare results by unit area. Finally, some of them also said that they charted student scores across assessments to monitor progress. For the off-the-shelf assessments, mathematics teachers said that they received reports that summarized results on student proficiency and need for additional instruction by standard; some assessments provided comparative data across homerooms and teachers; other assessments provided daily, weekly, and monthly reports on student performance broken down by clusters of the common core standards; and finally, some assessments also provided item-level performance data.

English Language Arts

Most of the ELA teachers who developed their own assessments said that they entered item-level student data into either Google forms or a spreadsheet and used it to create charts, analyze understanding by question-type, and track overall class mastery level. For the off-the-shelf assessments, ELA teachers said that they typically received class rosters and student reports that were disaggregated by demographic information, and for some assessments, they obtained results that compared student performance on previous assessments and predicted their targeted growth for the next assessment. They also indicated that they obtained results at the skill level (e.g., phonics, word recognition, and spelling) for some of the off-the-shelf unit assessments.

Use Cases

In order to understand current practices with respect to use of preunit versus end-of-unit assessments, we asked teachers how often they used unit assessments at the beginning of each curriculum unit as compared to at the end of that unit. Overall, across all focus groups, both mathematics and ELA teachers indicated that they rarely used beginning-of-unit assessments. Use of end-of-unit assessments was more common for both content areas, but more so for mathematics than for ELA. ELA teachers mentioned using unit assessments for writing and for grammar, but they did not use unit assessments for reading. For writing, teachers used preunit assessments almost every time a new genre of writing (e.g., narrative, informational, persuasive) was introduced; teachers are specifically interested in preunit (typically assessed using a quick-write task) to end-of-unit (typically assessed through projects assigned to create a written piece) assessment of growth for each writing genre. For grammar, preunit assessments are again commonly used. These assessments are most useful to teachers when they break down student performance by diagnostics (e.g., parts of speech) around areas with which students are struggling the most. However, reading is not structured in a unit format for a number of elementary ELA classrooms; therefore, no unit assessments are used.

One of the main reasons why preunit assessments were not frequently used in both mathematics and ELA was time constraints. Teachers said that they did not currently have the time to set aside one class period for preunit assessment and one for end-of-unit assessment for each unit covered during the year; however, many teachers indicated that they used formative quizzes to assess students' understanding as they move through the unit. All teachers said that they would be very interested in preunit assessments, particularly those that targeted prior knowledge before the start of a new unit, but they would only be able to realistically administer them for each unit if these assessments were short (not even 45 min). The time constraint factor in administering preunit assessments, despite teachers' needs to understand students' prior knowledge, is clearly articulated by one ELA teacher:

> "We do more informal beginning of unit assessments – I don't want to waste my time doing a full-fledged assessment – I just want to get an idea of where they stand … Throughout the lesson, I will do mini-checks since it is important for me to know if the students understand what I am teaching. I then follow-up with an end-of-unit assessment – and evaluate if they need any additional resources or additional help for this curriculum unit."

The time constraint factor mentioned by teachers has implications for the design of reports for unit assessments. Reports that indicate progress from preunit to mid-unit to end-of-unit should be designed in a modular format such that they show results for the tests actually administered. If only one test was administered for a unit, the interface should be designed so that it can be adapted to just show those results and not show progress within an unit. On the other hand, if multiple assessments are administered for a given unit, then progress within the unit should be displayed.

Top Needs for Results From Unit Assessments

During the first whole-group discussion, we asked teachers to identify the types of feedback they would value from unit assessments in order to begin to understand teachers' needs for results from the unit assessments. This discussion built on the types of assessments they currently used and types of feedback they reported receiving in the premeeting task. At the end of the first whole-group discussion, teachers were asked to reflect upon the preceding discussion and individually identify their top five needs for results from unit assessments. Teacher responses across all six focus groups were synthesized and high-level categories of needs were identified across these responses. As we analyzed the top needs indicated by each teacher, we identified some high-level categories or patterns for the types of needs that emerged. These high-level categories along with the counts (number of times these needs were mentioned across the 23 teachers) are presented separately in Table 2 for the mathematics and ELA teachers.

Mathematics

Table 2 shows the eight unique need categories that emerged for the mathematics teachers. Of these, the need to identify students' level of mastery or current level of understanding either based on the standards or based on skills was identified by all mathematics teachers in our sample. Teachers' specific comments indicated that they would like to see students' level of understanding and their attainment of mastery by curricular units presented as leveled groups for the whole classroom.

Table 2 Top Need Categories for Mathematics and English Language Arts (ELA) Teachers

Category	Mathematics ($n = 12$)	ELA ($n = 11$)
Identify students' level of mastery based on standards/skills	12	4
Student and parent friendly reports with visual aids	11	10
Student strengths and weaknesses with suggestions/resources to bridge the gap in understanding	10	11
User friendly features (visual, navigation, and data aggregation)	7	5
Grouping information based on needs and performance	5	7
Track students' progress	5	
Item level analysis		4
Whole class performance analysis for instruction	2	4
Compare and share results with other teachers and administrators	4	

A number of teachers also mentioned being able to see this information broken down by the standards (Common Core or state standards). In addition, some mathematics teachers specifically mentioned LP levels as a means for organizing and presenting student performance on unit assessments. For example, one teacher said: "I would like the results to specifically unpack the standards for remediation: what specific elements of the standard was/was not met to result in the level on the learning progression." It is important to note here that this type of request for an LP-based organizational framework was elicited even before we showed them our current mock-ups.

The next most popular need reported by the mathematics teachers was to have printable student and parent-friendly reports with appropriate visual aids (and appropriate nontechnical language) that can help teachers share information either with individual students to aid them in understanding their own strengths and weaknesses or with their parents during a parent–teacher conference. In addition, all teachers mentioned that these parent-friendly reports should also provide information that parents can use at home with their child. Moreover, some teachers also mentioned the need to share results at the individual student and/or class level with administrators or other teachers (e.g., ESL teachers) who work with those students. Finally, several teachers pointed out the importance of students having the ability to see and understand their own results, which gives them some accountability for their own learning.

Other top needs that emerged for mathematics teachers include obtaining a summary of student strengths and weaknesses at the class level and at the individual student level, user-friendly features (visual, navigation, and data aggregation), grouping information based on performance, tracking students' progress, whole class performance analysis, and the ability to compare and share results with other teachers (see Table 2 for the full list of top needs for mathematics teachers).

English Language Arts

Table 2 shows the seven unique need categories that emerged for the ELA teachers. Of these, obtaining feedback on student and class-level strengths and weaknesses with targeted suggestions and resources for next steps emerged as a top need for ELA teachers. Teachers were particularly interested in identifying gaps at the individual student level so that targeted remediation and instruction could be provided to individual students. Teachers also indicated that it was important that these strengths and weaknesses be aligned to corresponding standards on unit assessment reports. ELA teachers were particularly interested in specific next steps and resources for remediation based on identified strengths and weaknesses. They wanted these recommendations to be targeted and individualized to each student and thought that "generic" statements at the group level would not be very helpful for successful remediation.

Similar to the mathematics teachers, having access to parent-friendly reports was a top need for the ELA teachers. ELA teachers also reiterated the importance of having these reports in student-/parent-friendly language to be shared at parent–teacher conferences. They wanted these reports to include meaningful visual aids to help students understand their strengths and areas where additional work is needed. For example, one teacher said: "Visual aids that put assessment results in student language is important so that students can begin to interpret their own data. Doing so will empower them to focus on self-improvement." They also thought that unit assessment reports would be very helpful as documentation in support of student assignment to response to intervention or accelerated learning groups.

Table 3 Overall Count of Likes, Dislikes, and Confusions for Mathematics and English Language Arts (ELA) Teachers by Report Screen

Figure	Likes	Dislikes	Confusions
Mathematics			
Figure A1: Class-level results for one assessment	29	3	4
Figure A2: Class-level narrative feedback for one assessment	21	2	1
Figure A3: Detailed feedback for one student	26	2	7
Figure A4: Class-level results for one LP (pre–post)	12	2	5
Figure A5: Class-level results across all LPs	19	2	8
ELA			
Figure A6: Class-level results for one assessment	21		11
Figure A7: Class-level narrative feedback for one assessment	15	7	6
Figure A8: Detailed feedback for one student	16	4	6
Figure A9: Class-level results for one KP (pre–post)	12	5	4
Figure A10: Class-level results across all KPs	12	4	10

Note. KP = key practice; LP = learning progression. See Appendix A for figures.

Table 4 Categories of Reactions to Preliminary Mock-Ups Disaggregated by Count of Likes, Dislikes, and Confusions for Mathematics and English Language Arts (ELA) Teachers

	Mathematics			ELA		
Category	Likes	Dislikes	Confusions	Likes	Dislikes	Confusions
Data included in reports	14		2	5	2	
Data presentation/navigation	16		3	20		2
Grouping	12	3	2	10		3
LP levels and descriptors	16	2	3	6	5	7
Next steps/resources	10		1	7	2	1
Other (color coding, sorting)	8	2	4	4	2	6
Progress	6		1	4	2	3
Sharing with parents/students	4	1	3	4	4	2
Standards	11	1	1	7	1	9
Test name, forms, and choice	4	2	5	2	1	1
Visualization	5			7	1	1

Note. LP = learning progression.

Other top needs that emerged for ELA teachers included grouping information based on needs and performance, user friendly features, identifying students' level of mastery based on standards/skills, item-level analysis, and whole-class performance analysis (see Table 2 for the full list of top needs for ELA teachers). ELA teachers expressed that it would be extremely useful for a report to include suggestions for student grouping information along with targeted next steps.

Evaluation of Preliminary Mock-Ups by Screen

Once all teachers had identified their top needs, we presented the preliminary mock-ups designed for these interim unit assessments (see Appendix A) to evaluate the extent to which these mock-ups were liked and understood by teachers. To accomplish this, we asked participants to mark up the report screens (using an online heat-map tool) and indicate what they liked, what they disliked, and what they found confusing on each screen. Participants clicked on various areas of the report screen and indicated if they liked or disliked something or found something confusing. The results from participant clicks were tabulated specifically for their likes, dislikes, and confusions, and their comments were then evaluated and classified into categories of responses. A high-level summary of teacher reactions to the mock-ups are presented in Tables 3 and 4. Table 3 shows the overall count of likes, dislikes, and confusions for the mathematics and ELA teachers by report screen.

As indicated above, participants clicks of likes, dislikes, and confusions were reviewed and categorized for each report screen to identify themes in participant preferences and interpretations from these preliminary mock-ups. Table 4 shows a summary of the categories of reactions to the set of preliminary mock-ups disaggregated by counts of likes, dislikes,

Table 5 Overall Count of Likes, Dislikes, and Confusions for the Screen Representing Class-Level Results for One Unit Assessment for Mathematics Teachers

Category	Likes	Dislikes	Confusions
Data included in reports	6		1
Data presentation/navigation	7		
Grouping	3	3	2
LP levels and descriptors	2		
Other (color coding, sorting)	4		1
Standards	5		
Test name, forms, and choice	1		
Total	29	3	4

Note. LP = learning progression.

and confusions for the mathematics and ELA teachers. Table 4 is only intended to present a high-level summary of the categories; while we hope that several of the category labels used in Table 4 are quite self-explanatory, we urge readers to look for elaboration within specific reporting screens and hope that specific comments about likes, dislikes, and confusions for each reporting screen may help clarify the category labels presented in Table 4. In the rest of this section, we will discuss mathematics and ELA teachers' reactions to each reporting screen.

Screen 1: Class-Level Results for One Unit Assessment

Mathematics

Overall, most mathematics teachers liked the data that were included and the way in which the data were presented in this screen (see Appendix A, Figure A1). Counts for the broad categories of comments made by mathematics teachers related to Screen 1 are summarized in Table 5.

Likes. The mathematics teachers in this study valued much of the data included in this screen. They made particular references to the inclusion of raw score data (i.e., percentage raw score) in this screen. For example, one teacher specifically said: "I really like that you get both a raw percentage score and a level score for all students in one class. I would find that very helpful." They also liked the roster view with the list of all students in the class. In addition, mathematics teachers were particularly impressed with the way in which data were presented on this screen; they liked being able to see all students in the class if they needed to make any quick comparisons between students. Several also mentioned that the ability to toggle between classes was a very useful feature. Teachers also liked that students were placed into LP levels and found this information very useful. A number made positive comments about the use of color to demarcate LP levels. Finally, teachers particularly liked that the related standards were listed on the top of the page. For example, one teacher said: "I really like that the standards are listed. This will allow me to use this data and integrate the results into the materials that I currently use."

Dislikes and Confusions. For the mathematics teachers, dislikes and confusions on this screen were all centered on the column representing groups. Two of the teachers were unsure what "groups" meant in this context; they wondered if these were current groups or suggested groups based on the results from this administration. One teacher was also confused by the date column and did not understand why students would take this test on different dates. Finally, one teacher was not sure if all columns were sortable and wanted a multisort feature where teachers could sort students not only by name but also by score and LP level.

English Language Arts

ELA teachers also liked the data that were included in this screen (see Appendix A, Figure A6) but were somewhat more confused by a number of data elements presented. Counts for the broad categories of comments made by ELA teachers related to the first screen are summarized in Table 6.

Likes. The ELA teachers in this study liked a lot of the data that were included in this screen and the way in which this data were presented. They really liked that both the raw score percentage and the student's KP level were included. In terms of data presentation and navigation, teachers were specifically impressed with the ability to easily separate and

Table 6 Overall Count of Likes, Dislikes, and Confusions for the Screen Representing Class-Level Results for One Unit Assessment for English Language Arts Teachers

Category	Likes	Dislikes	Confusions
Data included in reports	3		
Data presentation/navigation	7		
Grouping	3		3
KP levels and descriptors			2
Other (color coding, sorting)	2		2
Standards	4		4
Visualization	2		
Total	21		11

Note. KP = key practice.

toggle through students by class period (or section). ELA teachers also really liked all the navigational tabs presented on the top of the screen to toggle from one report view to the next. They also liked the roster view, which helped them to get a quick overview of class performance and make generalized observations. ELA teachers also liked the color-coded levels and thought that the visualization was much more user friendly and meaningful than many other reports that they had viewed. Finally, ELA teachers really liked that the results for unit assessments were tied to specific standards and that they could easily see the relevant standards on this screen.

Dislikes and Confusions. As can be seen from Table 6, the ELA teachers in this study did not really dislike anything about this report screen. However, ELA teachers were somewhat confused by a number of data elements presented on this screen. Again, similar to the mathematics teachers, ELA teachers were unsure what "groups" meant in this context and how students were assigned into groups for the data presented in Column 2; they wondered if these groups corresponded to the levels (i.e., KP levels) presented in the last column and thought that this was redundant information if these two columns were correlated. ELA teachers were also generally confused by the KP levels and thought that these corresponded to performance levels; they wondered if the cut scores were for getting placed into these levels and where they could find that information. One teacher assumed that the KP levels were grades and said the following: "This appears to simply be turning a traditional percentage into a 1, 2, 3, 4, standard grade that is not what standards-based grading is really supposed to be." Teachers were also confused if all columns were sortable or if only the names were sortable. Similar to the mathematics teachers, they also wanted a multisort feature. Finally, ELA teachers were confused by the standards presented at the top of the screen; they were unsure if that section corresponded to the Common Core State Standards or their state standards of if it could be customized for each state. They also wondered if the standards were hyperlinked and if clicking on them would take them to the full description of the standards, because standards presented as an alpha-numeric code were not very meaningful to them.

Screen-2: Class-Level Narrative Feedback for One Unit Assessment

Mathematics

This screen (see Appendix A, Figure A2) was a favorite for the mathematics teachers; a number of teachers reacted very positively to the student grouping and LP level narrative feedback provided in this screen. Counts for the broad categories of comments made by mathematics teachers related to this screen are summarized in Table 7.

Likes. Mathematics teachers in this study really liked the suggested grouping information provided on this report screen. The most prominent feature they found attractive on this screen was the narrative feedback provided within LP levels. For example, one teacher said: "I love the descriptors! It helps me understand where a student is truly struggling, and allows the teacher to pinpoint the trouble areas." They also liked that the LP levels were further broken down based on students' specific gaps in understanding. They liked that students were grouped based on not only their LP level but also their specific weaknesses within a particular LP level. They thought that this feedback would be invaluable in saving them time and in helping them focus on immediate remediation and targeted next steps for each group of students. Finally, two mathematics teachers also liked that the results were linked to standards on this page.

Table 7 Overall Count of Likes, Dislikes, and Confusions for the Screen Representing Class-Level Narrative Feedback for One Unit Assessment for Mathematics Teachers

Category	Likes	Dislikes	Confusions
Data presentation/navigation			1
Grouping	6		
LP levels and descriptors	10	1	
Next steps/resources	1		
Other (color coding, sorting)	1		
Sharing with parents/students	1		
Standards	2		
Test name, forms, and choice		1	
Total	21	2	1

Note. LP = learning progressions.

Table 8 Overall Count of Likes, Dislikes, and Confusions for the Screen Representing Class-Level Narrative Feedback for One Unit Assessment for English Language Arts Teacher

Category	Likes	Dislikes	Confusions
Data presentation/navigation	1		
Grouping	7		
KP levels and descriptors	5	4	2
Next steps/resources		1	
Other (color coding, sorting)	1	1	1
Sharing with parents/students		1	1
Standards	1		2
Total	15	7	6

Note. KP = key practices.

Dislikes and Confusions. Overall, very few things were either disliked by or caused confusion for mathematics teachers on this screen (see Table 7). One teacher made two comments about the choice of labels (i.e., she did not like the term "descriptors"). And, one teacher mentioned that the purple dots (indicated on some class periods but not on others) were unclear and confusing.

English Language Arts

ELA teachers also liked the grouping information provided on this screen (see Appendix A, Figure A7), but unlike the mathematics teachers, ELA teachers did not all like the narrative feedback provided within KP levels on this screen. Counts for the broad categories of comments made by ELA teachers related to this screen are summarized in Table 8.

Likes. Similar to the mathematics teachers, ELA teachers also loved the grouping information presented in this screen. For example, one teacher said: "I LOVE THIS!!! This hits my needs as a teacher. Not only does it group students in small groups, it identifies areas for instructional growth for each group. This is exactly what I need!" Several teachers commented on the benefit of this feedback for their small-group instructional needs. A few ELA teachers also liked the KP level descriptors and the narrative descriptions provided for teachers.

Dislikes and Confusions. ELA teachers disliked some aspects of the narrative feedback provided on this screen. Specifically, teachers did not like the "not yet Level 2" label, and a few of them disliked these narrative descriptions and thought that they were very generic. One teacher also thought that the directions for next steps provided within this feedback was somewhat ambiguous for students grouped into different buckets. For example, this teacher said:

> This gives a specific area of concern, however, does not provide an example of the student error. Phrases such as, "generally able" or "may need support" are only "kind of" helpful. So, the teacher needs to read the questions that were incorrect to determine where point of instruction needs to occur."

Table 9 Overall Count of Likes, Dislikes and Confusions for the Screen Representing Detailed Feedback for one Student for the Mathematics Teachers

Category	Likes	Dislikes	Confusions
Data included in reports	6		
Data presentation/navigation	4		
Grouping	1		
LP levels and descriptors	2		2
Next steps/resources	8		1
Other (color coding, sorting)	1		
Progress	1	1	
Sharing with parents/students	1	1	3
Standards	2		1
Total	26	2	7

Note. LP = learning progressions.

In addition, one teacher thought that the language used in this screen was not parent friendly to share with parents (although not the intent of this screen), and two teachers did not understand how to access the standards relevant to KP levels presented on this screen.

Overall, there were some broad misunderstandings about the KPs among the ELA teachers. In particular, some statements and comments made with respect to the KP levels demonstrated a lack of familiarity with the KP framework among these teachers. For example, one teacher said:

> "I understand that the reports are designed with the assumption that the majority of students will fall in the middle levels. Additionally, one can probably assume that students scoring at Level 1 need help in just about everything covered during the assessment in order to show growth. However, it would be nice to know a little bit more about how students in the HIGHER PERFORMANCE LEVELS did. We know that, overall, they must have done well … but in what ways?"

Screen 3: Detailed Feedback for One Student

Mathematics

Overall, mathematics teachers strongly liked and/or endorsed the targeted next steps and resources provided on this screen for each individual student (see Appendix A, Figure A3), but they also presented some dislikes and confusions, particularly with regard to the ability to share information on this screen with parents. Counts for the broad categories of comments made by mathematics teachers related to this screen are summarized in Table 9.

Likes. Mathematics teachers liked a number of elements included on this screen; they specifically liked the raw score (e.g., 17 out of 22) information provided for each unit assessment for the student. They also liked that they could look at an individual student's performance across all units (LPs) and thought that this screen almost represented an individual data folder that they would maintain for each student in their class. Teachers also liked the "expand/collapse" feature and the ability to navigate from student to student within this screen. One teacher indicated that this screen would be very useful to share as an individual student progress report with parents, and two teachers were very positive about the connection to standards consistently presented on all report screens, including this screen.

Overall, the student-specific narrative feedback with suggested next steps and links to OERs was most attractive to mathematics teachers on this screen. Teachers thought that this would be a great way to help students and their parents find resources that they could use at home to better help themselves. For example, one teacher said: "I like that there are external links that can help the student. It helps to make it a little more parent-friendly, so a teacher can say "here is xyz [sic] that will help a student continue to master xyz [sic] concept." However, from most of the comments made about the links to OERs presented on this screen, it seems that teachers were looking at this as a resource to share with parents and not really as a resource to use in their differentiated instruction.

Dislikes and Confusions. As can be seen from Table 9, there were a few things that were confusing to mathematics teachers on this screen. Two teachers thought that the LP descriptor/narrative feedback that was provided was probably too

Table 10 Overall Count of Likes, Dislikes, and Confusions for the Screen Representing Detailed Feedback for One Student for English Language Art Teachers

Category	Likes	Dislikes	Confusions
Data presentation/navigation	3		
KP levels and descriptors	1		1
Next steps/resources	7	1	1
Other (color coding, sorting)			2
Progress		1	1
Sharing with parents/students	2	2	
Standards	2		1
Visualization	1		
Total	16	4	6

Note. KP = key practices.

wordy; they suggested putting a brief summary on this page and perhaps providing a link to a full explanation elsewhere. One teacher indicated that they would like to be able to see all assessments for one LP (across years) on the same screen so that they could monitor student progress on each LP. One mathematics teacher also wondered how these resources would be added and if there was a way for districts to add resources based on what was recommended or available within that district. Finally, three mathematics teachers were also confused if this page was intended to be parent and/or student facing or only teacher facing. They said that it would be nice if this page could be designed as a parent-facing/parent-friendly page so that they could just print this page and send it to parents or assign the links as homework for students. If the page were made sharable with the students and/or their parents, mathematics teachers indicated that the descriptions and narrative feedback on this screen would definitely have to be substantially shortened.

English Language Arts

ELA teachers, in general, also liked the feedback about next steps and resources provided on this screen (see Appendix A, Figure A8). Counts for the broad categories of comments made by ELA teachers related to Screen 3 are summarized in Table 10.

Likes. ELA teachers liked the ability to navigate between students on this screen. They also really liked how the data were presented visually (graphically) and how the KP levels were laid out at the student level. One teacher specifically liked the KP level description that helped them understand what students performing at that level can do. Several teachers mentioned that they really liked the suggested next steps and targeted resources that were provided at the student level; they particularly liked that targeted links would be provided to online resources to help students work on their specific areas of need. For example, one teacher said: "I like how this report gives clear next steps and the links and resources to work on those steps. Too often we get data, but no clear direction on how to move the student to the next level. This gives that next step, which frees my time as a teacher to implement it." Two teachers also thought the narrative feedback would be very helpful for a teacher, especially to share with parents during the parent–teacher conference or to write within student report cards.

Dislikes and Confusions. There were a few elements on this screen that ELA teachers disliked and/or found confusing. One of the teachers thought that the next steps were weak even with the links to OERs. Similar to what was found among mathematics teachers, one ELA teacher indicated that it would be helpful to see progress within one KP on this individual student screen. This idea was also mentioned during the whole-group discussion that followed, and many ELA teachers mentioned that it would be useful to be able to see preunit and end-of-unit results for each KP on the individual student screen. Two ELA teachers mentioned that it would be useful if this screen were more student friendly. For example, one teacher said:

> "I don't think this page is very student-friendly. This works for me, but it certainly wouldn't help a student understand where they need to go. It's too wordy for them, and too much teacher-speak. This applies also down at the bottom where there's a great paragraph for ME about what Chandra can and can't do, but now I have to translate this for her."

Table 11 Overall Count of Likes, Dislikes, and Confusions for the Screen Representing Class-Level Results for One Learning Progression (Pre – Post) for Mathematics Teachers

Category	Likes	Dislikes	Confusions
Data included in reports			1
Data presentation/navigation	1		
LP levels and descriptors	2		
Next steps/resources			
Other (color coding, sorting)			
Progress	6		1
Sharing with parents/students	1		
Standards		1	
Test name, forms, and choice	2	1	3
Total	12	2	5

Note. LP = learning progression.

During discussions, ELA teachers also indicated that if this screen is printed to share with parents or students, they would like the ability to hide other student names from this screen. One of the ELA teachers said that they would like results presented in terms of standards (rather than as LP/KP levels) with an in-depth breakdown of what each standard represents and resources for remediation. Finally, with regard to the OERs, one teacher said the following:

> "While I love the thought of getting online resources, I am wondering if there is a way to build a data bank in your system that allows teachers to specify their state (and, therefore, which standards they use), and resources can then be provided that support that specific set of standards."

Screen-4: Class-Level Results for One LP (Pre–Post)

Mathematics

Overall, mathematics teachers liked the ability to see progress for all students in their class from preunit to mid-unit performance for a given LP on this screen (see Appendix A, Figure A4). Counts for the broad categories of comments made by mathematics teachers related to this screen are summarized in Table 11.

Likes. Mathematics teachers liked the side-by-side comparison of preunit to mid-unit performance for all students in their class, which helps them to see if their students improved, to intervene as necessary for individual students before the unit is completed, and to restructure lessons as necessary if the majority of students are not showing progress. Two mathematics teachers liked that different unit assessments were available for preunit and mid-unit assessment and that these assessment names (and pertinent levels) were clearly identified for each administration.

Dislikes and Confusions. As can be seen from Table 11, there were very few elements that mathematics teachers disliked on this screen. One teacher wondered if the raw score provided could be translated to a grade to help them with report cards. One teacher found it difficult to tell if a student made progress or not, but this issue was not prevalent. One teacher indicated that they wanted more details about the standards — not just what standards correspond to these LP, but more specific details about the standard and how the standards are addressed within this LP. Three teachers indicated that they were confused by the unit assessment forms and the number of assessments a student could have taken for each LP; they wondered if the student only ever took two assessments on linear functions or if they could see all of the results if the student took more than two assessments (from the current or previous years) on this screen. Finally, one teacher wanted to have the ability to group students based on which unit assessment they took.

English Language Arts

ELA teachers also liked the ability to see progress from preunit to mid-unit performance for all students in their class on this screen (see Appendix A, Figure A9). Counts for the broad categories of comments made by ELA teachers related to this screen are summarized in Table 12.

Table 12 Overall Count of Likes, Dislikes, and Confusions for the Screen Representing Class-Level Results for One Learning Progression (Pre–Post) for English Language Arts Teachers

Category	Likes	Dislikes	Confusions
Data included in reports	2	1	
Data presentation/navigation	2		1
Other (color coding, sorting)	1		
Progress	3	1	1
Sharing with parents/students	1	1	
Standards		1	1
Test name, forms, and choice	1	1	1
Visualization	2		
Total	12	5	4

Likes. Overall, ELA teachers liked the way in which data were presented in this screen and, specifically, that each unit assessment was named. They also liked how the levels were color coded, and they also liked the "most recent" feature, which would allow them to quickly access and see results for the most recent administration. They also liked the visual representations on this screen and how the scores for preunit and mid-unit were reported side by side for each student, allowing teachers to easily make comparisons about progress in a snapshot for individual students and their entire classroom.

Dislikes and Confusions. As can be seen from Table 12, ELA teachers also disliked a few elements presented on this screen and were confused by some of the data presented on this screen. One ELA teacher was confused by the KP language and wondered if these results could be organized by state standards. One teacher was confused if the student would be taking the same unit assessment again or if there was a different form/version for the same level. One teacher wanted to see specific growth information that could be linked to each individual student so that their progress on the KP could be monitored. One teacher was confused by the presentation of raw score when all students did not take the same unit assessment ("It is like comparing apples and oranges."). This teacher is not wrong, and it is important to consider this input in the redesign of reports. During discussions, several ELA teachers reiterated again that they wanted to be able to link the information presented here to a student page so that they could see progress from preunit to mid-unit to end of unit for a given individual student across the entire spectrum of the KP. One ELA teacher wanted to get rid of the KPs and display results in terms of "true standards-based grading." Finally, one ELA teacher had concerns about choosing their own unit assessment, noting: "Many districts may want students to complete the same assessment and may not allow teachers to have the ability to choose which assessment to administer."

Overall, taking the comments made across the last couple of screens together, it can be concluded that access to the entire history of all unit assessments administered for a given LP/KP at the individual student level seems very important for both mathematics and ELA teachers.

Screen-5: Class-Level Results across all Learning Progressions

Mathematics

Overall, mathematics teachers liked the way in which data were presented and particularly liked the visualization representing class performance across LPs on this screen (see Appendix A, Figure A5), but they displayed relatively more confusions about information presented on this screen compared to other screens. Counts for the broad categories of comments made by mathematics teachers related to this screen are summarized in Table 13.

Likes. Many mathematics teachers liked the data that were included on this screen and the way in which data were presented across all LPs on this screen; they particularly liked the graphical representations (including the way in which the bars were color coded) showing results across LPs. Two teachers indicated that this screen provided them with a quick way to compare and group various students. For example, one mathematics teacher said: "I love being able to see where the class is as a whole with each topic. This is very helpful, especially when topics build on top of each other, to be able to see which students are struggling and where." One mathematics teacher thought that this screen would be a good resource for them to share with other teachers or administrators.

Table 13 Overall Count of Likes, Dislikes, and Confusions for the Screen Representing Class-Level Results Across All Learning Progressions (LPs) for Mathematics Teachers

Category	Likes	Dislikes	Confusions
Data included in reports	2		
Data presentation/navigation	3		2
Grouping	2		
LP levels and descriptors		1	1
Other (color coding, sorting)	2	1	3
Progress	1		
Sharing with administrators	1		
Standards	2		
Test name, forms, and choice	1		2
Visualization	5		
Total	19	2	8

Table 14 Overall Count of Likes, Dislikes, and Confusions for the Screen Representing Class-Level Results Across All Key Practices (KPs) for English Language Arts Teachers

Category	Likes	Dislikes	Confusions
Data included in reports		1	
Data presentation/navigation	7		1
KP levels and descriptors		1	4
Other (color coding, sorting)		1	1
Progress	1		1
Sharing with parents/students	1		1
Standards			1
Test name, forms, and choice	1		
Visualization	2	1	1
Total	12	4	10

Dislikes and Confusions. The representations on this screen resulted in a few confusions for mathematics teachers. Specifically, mathematics teachers indicated that they wanted to be able to compare their whole-group data with other homerooms (their own or other teachers' so they can facilitate data-based discussions) easily on the same screen. If that data cannot be provided on this screen, they wanted to be able to see another screen where they could easily make those comparisons. One teacher specifically indicated that it would be important to be able to easily see the list of students at each LP level and wondered if hovering over each color on the stacked bars would show them the specific students at that level for the given LP. One teacher was confused about where and how they would find the specific level descriptors for each LP and asked: "Where can I go to find what Level 5 on Proportional Reasoning, for example, means?" Two teachers were unsure why mid-unit results were presented for some LPs and end of unit results were presented for others; they wanted to be able to see preunit and end-of-unit results or toggle between views where they can see preunit and end-of-unit results on this screen. Finally, three mathematics teachers indicated they wanted the information (i.e., LP levels) provided in the student roster/table in the bottom half of the screen to be color coded for levels (and sortable by levels) to match the rest of the screens in the report.

English Language Arts

ELA teachers also generally liked the way in which data were presented on this screen (see Appendix A, Figure A10), but they also presented some significant confusions. Counts for the broad categories of comments made by ELA teachers related to this screen are summarized in Table 14.

Likes. Several ELA teachers liked the way in which data were visually represented on this screen. They felt that this representation where data were combined across all KPs was very useful and easily gave them a "big picture" view of their class. They liked that they could just look at results for what they have covered so far that year. They liked the hyperlinks for student names and the ability to be able to click on a student name to directly go into the student's report. Teachers also

liked the ability to view the graphs and the students' KP levels listed in the roster view on the bottom half of the screen. Finally, one ELA teacher thought that it was very useful that these unit assessments were teacher initiated and that they would be able to administer an assessment if and when they needed (either before starting a unit, during, or at the end of a unit).

Dislikes and Confusions. ELA teachers were confused about several elements presented on this screen; a lot of this confusion was around the KP levels and the way in which they were represented on this screen. However, there were dislikes and confusions regarding data presentation and navigation as well. For example, similar to the mathematics teachers, ELA teachers also wondered if the bars on the chart were clickable and if they could access preunit and mid-unit results together on the same page in order to compare. Overall, however, as indicated earlier, most of the confusions for ELA teachers resulted from their unfamiliarity with the KP framework. It was confusing for several teachers that the entire classroom of students were distributed within Level 1 and Level 2 for one KP (e.g., Discuss and Debate Ideas), while there were students performing at Level 5 for other KPs. It was not clear to the teachers that some of these KPs could have been introduced at earlier grades and others could have been just introduced at the current grade. When clarified during discussions, teachers indicated that it should be easier to identify which skills were more newly introduced and which ones had been introduced at earlier grades.

In addition, there were other confusions surrounding the KP levels and the use-case identification of preunit, mid-unit, and end of unit for the most recent test administered. For example, one teacher said:

> "I am not sure that I understand what this report is showing me. It appears to be filtered to show all Key Practices. This leaves me to assume that the blank columns indicate concepts that have not yet been assessed. Assuming that is the case, why does the rest of the chart show Preunit, Mid-Unit, and End of Unit? If all three assessments are being displayed, when would the other two Key Practices be assessed? Am I to assume that this means those concepts were simply not covered in this specific unit?"

During discussions, it was apparent that ELA teachers were confused with KPs in general and would prefer results presented in terms of skills or standards that were being covered. For example, one teacher said: "This is confusing in that the scores are not attached to skills, just unit titles. It would be helpful to have a page that listed individual skills with each student's scores in order to complete report cards."

Missing Features and Suggestions for Improvement

Finally, after evaluating each of the screens independently, teachers were asked to evaluate the set of screens as a whole and consider if their top five needs (elicited prior to showing them the mock-ups) were addressed within these set of screens as a whole. In this section of the report, we summarize the specific needs that both mathematics and ELA teachers thought were missing from the current set of mock-ups. The overall categories of missing needs that emerged across both mathematics and ELA teachers are summarized in Table 15.

Table 15 shows that the need for a parent-/student-friendly report was something that both mathematics and ELA teachers thought was very important but missing from the current set of mock-ups. Readers should recall that this need ranked high for both mathematics and ELA teachers when top needs were elicited and was highlighted as a concern by both groups of teachers on the individual student report screen. Teachers wanted to have access to individual student results in a format that could be easily downloaded (printed) to share with parents and students and/or emailed to parents. Both mathematics and ELA teachers thought that it was very important to be able to share unit assessment results with parents during parent–teacher conferences. Specifically, teachers wanted to be able to assign homework for students in areas of need with links to online resources that the students could access and the parents could monitor. Therefore, both mathematics and ELA teachers thought that it was important to have the next steps and resources available in parent- and student-friendly language so that additional follow-up work may be assigned for students.

ELA teachers also thought that the narrative feedback on strengths and weaknesses (on Screen 2, see Figure A2) was too generic and should be individualized to the students and also presented in student-friendly language. Another prominent missing need that was mentioned by several ELA teachers, but not as many mathematics teachers, was the ability to get question-by-question reports on what students got correct versus incorrect. Finally, ELA teachers also thought that some next steps for the class as a whole should be provided along with class average scores.

Table 15 Missing Needs for Unit Assessment Results From the Current Mock-Ups

Missing needs	Mathematics	ELA
Parent-/student-friendly reports	6	9
Next steps linked to resources that also help parents/students	3	3
Compare across classrooms and teachers and share with other faculty	3	2
View correct/incorrect responses by question to inform instruction	1	5
Overall snapshot of strengths and weaknesses for the whole class with class average and next steps for class as a whole	1	3
Ability to group and sort students based on level	3	1
More specific feedback on strengths and weaknesses of students – language in narrative is too generic		3
Full text of standards	2	
Disaggregated based on demographics	1	
Progress since previous unit assessment		2
View prerequisite knowledge for an LP	1	
Color-code beginning of unit/mid-unit/end of unit	1	
Add district-specific resources	1	

Note. ELA = English language arts; LP = learning progression.

Table 16 Mathematics Teachers' Familiarity and Interest in Five Popular Open Educational Resources (OERs)

OER	Heard of the resource	Used the resource	Recommend — based on use	Interested — based on the description	Would like to have it included
Khan Academy	8 (100%)	8 (100%)	8 (100%)		8 (100%)
OpenEd	2 (25%)			8 (100%)	6 (75%)
OER Commons				8 (100%)	3 (38%)
Curriki	1 (13%)	1 (13%)	1 (13%)	6 (75%)	5 (63%)
Education.com	3 (38%)	2 (25%)	1 (13%)	6 (75%)	5 (63%)

Note. The question about OERs was not included in the premeeting assignment for the first face-to-face focus group; therefore, the total number of mathematics teachers who responded to these questions only includes those in the two virtual focus groups conducted for mathematics (i.e., $n = 8$).

Table 17 Other Useful Open Educational Resources (OERs) Mentioned by Mathematics Teachers

OER	Count	Notes
Study Island	4	Less game-like and more content-based, but, also perhaps a paid service
Learn Zillion	4	Considered more user-friendly than Khan Academy, more of a teaching tool, not really game-based
Prodigy	4	Provides Math games and useful reports by skill-level and usage
Other resources mentioned	1	ST-Math//Moby Max math//Amplify//EdPuzzle//Gilder Lehman Institute//Teachers pay teachers//Engage NY//Sum Dog//Quizizz

Open Educational Resources

For the five virtual focus groups, participants were asked about their familiarity with five popular OERs (Khan Academy, OpenEd, OER commons, Curriki, and Education.com) during the premeeting assignment. Before concluding each of these focus groups, we returned to the discussion around OERs. We asked participants to indicate not only their familiarity with the five OERs listed, but the extent to which the OERs listed in the premeeting were particularly useful in their context. In addition, they were asked to identify any additional OERs (beyond the five identified in the premeeting assignment) that they may have used and describe what they found useful about these OERs. Participant responses to the familiarity with the five popular OERs as reported during the premeeting assignment are summarized in Tables 16 and 17.

Table 18 English Language Arts Teachers' Familiarity and Interest in Five Popular Open Educational Resources (OERs)

OER	Heard of the resource	Used the resource	Recommend based on use	Interested based on the description	Would like to included
Khan Academy	11 (100%)	7 (64%)	7 (64%)	3 (27%)	10 (91%)
OpenEd	3 (27%)			9 (82%)	8 (73%)
OER Commons				9 (82%)	3 (27%)
Curriki	3 (27%)			8 (73%)	4 (36%)
Education.com	9 (82%)	7 (64%)	4 (36%)	3 (27%)	4 (36%)

Note. A dash indicates that responses were not obtained.

Table 19 Other Useful Open Educational Resources (OERs) Mentioned by English Language Arts Teachers

OER	Count	Notes
Teachers Pay Teachers	3	Free to use, but many resources are paid—great for ELA
Study Island	2	Less game-like, but, also perhaps a paid service
Noredink.com	2	Great writing resource; helps students with grammar; also provides great reports
ereadingworksheets.com	2	completely FREE—provides worksheets for working with specific kids
Other resources mentioned	1	CommonLit//Discovery Ed//LearnZillion//Readworks//Engage NY//Simple K12//ReadWriteThink//Google Education Suite//links to YouTube explanations of topics//Florida Center for Reading Research//MyON//Flocabulary//IXL.com

Based on the premeeting responses presented in Table 16, it can be seen that all the math teachers (i.e., only those in the two virtual focus groups) had both heard of Khan Academy and used it and would like this resource if it were included in the unit assessment reports. Mathematics teachers in this study were not as familiar with the other OERs included in this list, but based on the descriptions provided in the survey, all teachers were interested in OpenEd and OER commons, and 75% (i.e., six of eight) teachers were interested in Curriki and Education.com. However, when asked if this was a resource they would like included within the unit assessment reports, only 38% (i.e., three of the eight teachers) indicated that they would like OER commons included in the reports.

As we returned to the discussions around OERs at the conclusion of the virtual focus groups, we specifically asked teachers if there were other OERs than the five listed during the premeeting assignment that they were familiar with and would recommend being included in reports. Table 17 summarizes the set of other OERs mentioned by mathematics and ELA teachers and their specific comments about what they liked or found useful within those resources. Table 17 notes that Study Island, Learn Zillion, and Prodigy were commonly mentioned by mathematics teachers. In addition, several other OERs (e.g., Moby Math, Sum Dog) were mentioned at least once by various mathematics teachers, and these have all been listed in Table 17.

Premeeting responses for the ELA teachers are presented in Table 18. It is apparent that the five OERs listed in the premeeting were not among the ones ELA teachers would like included in the test reports. Again, all of the ELA teachers had heard about Khan Academy; several of them (7 of 11) had used it and did indicate that they would like this resource included in the reports. However, it became apparent during the focus group discussions that Khan Academy did not have any specific ELA content (this was true at the time these focus groups were conducted), and a number of the elementary ELA teachers thought Khan Academy would be useful as a resource for other content (e.g., mathematics) but did not think that it would be a very useful for ELA.

When we returned to the discussions around OERs at the end of the focus groups, ELA teachers mentioned several OERs that were commonly used for ELA. These results are summarized in Table 19. More than one ELA teacher mentioned using Teachers Pay Teachers, Study Island, and Noredink.com; several other open and curated online resources were mentioned by ELA teachers (see Table 19).

Overall, the specific OERs that would be most useful to mathematics and ELA teachers in the unit assessment context are not clear from these results and additional follow-up surveys are warranted to determine which specific OERs teachers would find useful and the way in which they would like information from these OERs incorporated into the unit assessment reports.

Discussion

The purpose of this study was to identify stakeholder-specific needs for results from interim-unit assessments. The primary score users for unit assessment results are classroom teachers; therefore, the focus of this preliminary needs assessment to develop audience-specific score reports was on teachers. In order to understand teachers' specific needs, we started with an analysis of the underlying assessment claims and designed some preliminary mock-ups as suggested in the iterative multi-step score report development framework (Hambleton & Zenisky, 2013; Tannenbaum et al., 2016; Zapata-Rivera, 2011a). We designed preliminary report mock-ups based on previous research findings that provide some insights into how teachers make sense of data (Mandinach & Gummer, 2016; Marshall & Drummond, 2006; Zapata-Rivera et al., 2010; Zapata-Rivera et al., 2016; Zwick et al., 2008) and other research in the formative assessment context (Black & Wiliam, 1998; Kulik & Kulik, 1988; Nicol & MacFarlane-Dick, 2006; Sadler, 1989) that provides some insights into teachers' needs in the interim/unit and formative assessment contexts.

The primary goal of this study was to evaluate these preliminary mock-ups with elementary and middle school mathematics and ELA teachers for clarity and comprehension. In addition, we also wanted to elicit any additional needs for results that are critical to teachers but completely missing from these preliminary mock-ups. Similar to findings from some previous studies (e.g., Shepard, 2018; Shute, 2008; Zapata-Rivera et al., 2007), we found that teachers were primarily interested in unit assessment feedback that was actionable and instructionally useful. Some of the primary needs identified by teachers included an identification of students' level of mastery by unit, students' specific student strengths and weaknesses, suggested small grouping information, and some targeted next steps. Teachers also highlighted the need to share unit assessment results with parents during parent – teacher conferences and with students so that they could better understand their own strengths and weaknesses.

In general, the LP/KP framework was used to present the unit assessment results was relatively more favored by the mathematics teachers than the ELA teachers. Overall, mathematics teachers made 21 different comments about the LP framework; of these, 16 were likes, two were dislikes, and three were confusions. ELA teachers made a total of 19 comments about the KP framework; of these, six were likes, six were dislikes, and seven were marked as confusions. Almost all the mathematics teachers in this study overwhelmingly liked the grouping information and the narrative feedback provided within LP levels on the second report screen (see Appendix A, Figure A2). They also liked that the LP levels were further broken down and that students were grouped based on their strengths and weaknesses within a particular LP level; teachers thought that this feedback would be invaluable for providing small group instruction. In contrast, ELA teachers were not as familiar with the KP framework and were confused by the grouping information presented within KP levels. Unlike the mathematics teachers, ELA teachers did not categorically like the narrative feedback provided within KP levels. Many ELA teachers reported that this information was too generic to be helpful, and similar to the previous findings (e.g., Zapata-Rivera et al., 2012), some teachers in our study found the narrative feedback to be somewhat overwhelming.

Overall, there were some general confusions regarding the KP levels and descriptors for the ELA teachers, and they mistook these KP levels for performance levels with associated cut scores. One reason for this general confusion about KPs could be a lack of familiarity with the KP framework. The KP framework was developed at ETS to help teachers recognize the interdependence of reading and writing skills and how they are integrated in order to address "big ideas" in ELA such as developing and presenting an argument. Most of the ELA teachers in our study were probably never instructed in this framework within teacher education programs or through ongoing professional development in their schools. Therefore, concerted efforts must be made, either by the testing company or the district/state, to provide targeted professional development to teachers in the LP/KP framework across all grades and across both mathematics and ELA.

To the extent that teachers understood the results presented in the LP/KP framework, they viewed it favorably. However, it was apparent from several comments made by both ELA and mathematics teachers that it was critical to present results across the entire spectrum of any given LP/KP; these teachers wanted to be able to see the entire history of all unit assessments administered for any given LP/KP for each student so that students' progressive understanding of the construct (domain) may be monitored.

Finally, one of the main needs for teachers that was not addressed in the current mock-ups was the need to have unit assessment reports in student–/parent-friendly language and format so it could be shared with students in discussions about their progress and with their parents during parent – teacher conferences. Teachers particularly thought that having

a student-facing report would be very useful to assign additional work on areas of need. They wanted these student-/parent-facing reports to be designed with simple language; to be printable; and, specifically, to include a summary about areas of need, next steps, and customized resources for the student.

The results from this study indicate that the preliminary mock-ups were viewed favorably by both mathematics and ELA teachers, and there were several features and data elements included in the current mock-ups that were liked by these teachers. However, some important features such as inclusion of parent- and student-friendly versions were considered missing, and the overall results presentation framework (i.e., within LP/KP levels) warranted some concerted efforts for teacher professional learning. The results from this study should be informative to various testing programs in the design and redesign of the score reports for interim unit assessments, and it is our hope that the results from this study are also informative to the field of score reporting in informing the types of information to include in teacher score reports for interim unit assessments.

References

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives*, *8*(2–3), 70–91. https://doi.org/10.1080/15366367.2010.508686

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–75. https://doi.org/10.1080/0969595980050102

Brown, G. T. L., O'Leary, T. M., & Hattie, J. A. C. (2019). Effective reporting for formative assessment: The asTTle case example. In D. Zapata-Rivera (Ed.), *Score reporting research and applications* (pp. 107–125). Routledge.

Charman, P. (2009, September 13–18). *Data rich, information poor: Creative and innovative approaches to results analysis to support teaching and learning* [Paper presentation]. International Association for Educational Assessment 35th Annual Conference, Brisbane, Australia. https://iaea.info/documents/data-rich-information-poor-creative-and-innovative-approaches-to-results-analysis-to-support-teaching-and-learning/

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence based approach to reform* (CPRE Research Report # RR-63). Consortium for Policy Research in Education. https://www.cpre.org/sites/default/files/researchreport/829_lpsciencerr63.pdf

Deane, P., Sabatini, J., Feng, G., Sparks, J., Song, Y., Fowles, M., O'Reilly, T., Jueds, K., Krovetz, R., & Foley, C. (2015). *Key practices in the English language arts (ELA): Linking learning theory, assessment, and instruction* (Research Report No. RR-15-17). ETS. https://doi.org/10.1002/ets2.12063

Goertz, M. E., Nabors-Olah, L., & Riggan, M. (2009, December). *Can interim assessments be used for instructional change?* (CPRE Policy Brief RB-51). Consortium for Policy Research in Education. https://doi.org/10.12698/cpre.2009.rb51

Goodwin, S. (1996). Data rich, information poor (DRIP) syndrome: Is there a treatment? *Radiology Management*, *18*(3), 45–49.

Greene, J. C. (1988). Communication of results and utilization in participatory program evaluation. *Evaluation and Program Planning*, *11*(4), 341–351. https://doi.org/10.1016/0149-7189(88)90047-X

Hambleton, R., & Zenisky, A. (2013). Reporting test scores in more meaningful ways: A research-based approach to score report design. *APA handbook of testing and assessment in psychology*. APA. https://doi.org/10.1037/14049-023

Hambleton, R. K., & Slater, S. C. (1997). *Are NAEP executive summary reports understandable to policy makers and educators?* (CSE Technical Report 430). National Center for Research on Evaluation, Standards and Student Teaching.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Routledge.

Hegarty, M. (2019). Advances in cognitive science and information visualization. In D. Zapata-Rivera (Ed.), *Score Reporting Research and Applications* (pp. 19–34). Routledge. https://doi.org/10.4324/9781351136501-3

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment*. Council of Chief State School Officers.

Impara, J. C., Divine, K. P., Bruce, F. A., Liverman, M. R., & Gay, A. (1991). Does interpretive test score information help teachers? *Educational Measurement: Issues and Practice*, *10*(4), 16–18. https://doi.org/10.1111/j.1745-3992.1991.tb00212.x

Kannan, P., Zapata-Rivera, D., & Leibowitz, E. A. (2018). Interpretation of score reports by diverse subgroups of parents. *Educational Assessment*, *23*(3), 173–194. https://doi.org/10.1080/10627197.2018.1477584

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*(1), 79–97. https://doi.org/10.3102/00346543058001079

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366–376. https://doi.org/10.1016/j.tate.2016.07.011

Marshall, B., & Drummond, M. J. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education*, *21*, 133 – 149. https://doi.org/10.1080/02671520600615638

Nicol, D. J., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies of Higher Education*, *31*(2), 199 – 218. https://doi.org/10.1080/03075070600572090

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. National Center for the Improvement of Educational Assessment. https://www.achieve.org/files/TheRoleofInterimAssessments.pdf

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119 – 144. https://doi.org/10.1007/BF00117714

Shepard, L. A. (2018). Learning progressions as tools for assessment and learning. *Applied Measurement in Education*, *31*(2), 165 – 174. https://doi.org/10.1080/08957347.2017.1408628

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153 – 189. https://doi.org/10.3102/0034654307313795

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic – molecular theory. *Measurement: Interdisciplinary Research and Perspectives*, *4*, 1 – 98. https://doi.org/10.1080/15366367.2006.9678570

Smith, J. P., III, Disessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, *3*(2), 115 – 163. https://doi.org/10.1207/s15327809jls0302_1

Tannenbaum, R. J., Kannan, P., Leibowitz, E. A., Choi, I., & Papageorgiou, S. (2016). *Interactive score reports: A strategic and systematic approach to development* [Paper presentation]. National Council on Measurement in Education Annual Meeting, Washington, DC.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Graphics Press.

Underwood, J. S., Zapata-Rivera, D., & VanWinkle, W. (2007). *Growing pains: Teachers using and learning to use IDMS* (Research Memorandum 08-07). ETS.

Wylie, E. C., & Lyon, C. J. (2017). Supporting teacher assessment literacy: A proposed sequence of learning. *Teachers College Record* (ID Number: 22193).

Zapata-Rivera, D. (2011a). Designing and evaluating score reports for particular audiences. In D. Zapata-Rivera & R. Zwick (Eds.), *Test score reporting: Perspectives from the ETS score reporting conference (Research Report No. RR-11-45)*. ETS. https://doi.org/10.1002/j.2333-8504.2011.tb02281.x

Zapata-Rivera, D. (2011b, April). *Designing score reports that help teachers make instructional decisions* [Paper presentation]. American Educational Research Association Conference, New Orleans, LA.

Zapata-Rivera, D., Hansen, E. G., Shute, V. J., Underwood, J. S., & Bauer, M. I. (2007). Evidence-based approach to interacting with open student models. *International Journal of Artificial Intelligence in Education*, *17*(3), 273 – 303.

Zapata-Rivera, D., & Katz, I. (2014). Keeping your audience in mind: Applying audience analysis to the design of score reports. *Assessment in Education: Principles, Policy & Practice*, *21*(4), 442 – 463. https://doi.org/10.1080/0969594X.2014.936357

Zapata-Rivera, D., VanWinkle, W., & Zwick, R. (2010, May). *Exploring effective communication and appropriate use of assessment results through teacher score reports* [Paper presentation]. National Council on Measurement in Education Annual Meeting, Denver, CO.

Zapata-Rivera, D., VanWinkle, W., & Zwick, R. (2012). *Applying score design principles in the design of score reports for CBAL™ teachers* (Research Memorandum No. RM-12-20). ETS.

Zapata-Rivera, D., Zwick, R., & Vezzu. (2016). Exploring the effectiveness of a measurement error tutorial in helping teachers understand score report results. *Educational Assessment*, *21*(3), 215 – 229. https://doi.org/10.1080/10627197.2016.1202110

Zwick, R., Sklar, J., Wakefield, G., Hamilton, C., Norman, A., & Folsom, D. (2008). Instructional tools in educational measurement and statistics (ITEMS) for school personnel: Evaluation of three web-based training modules. *Educational Measurement: Issues and Practice*, *27*, 14 – 27. https://doi.org/10.1111/j.1745-3992.2008.00119.x

Appendix A

Mathematics and English Language Arts Mock-Up Examples

Mathematics Mock-Ups Designed for a Grade 8 Unit Assessment (Going to Ceres: A Level 2–3 Unit assessment on the Learning Progression: Linear Functions

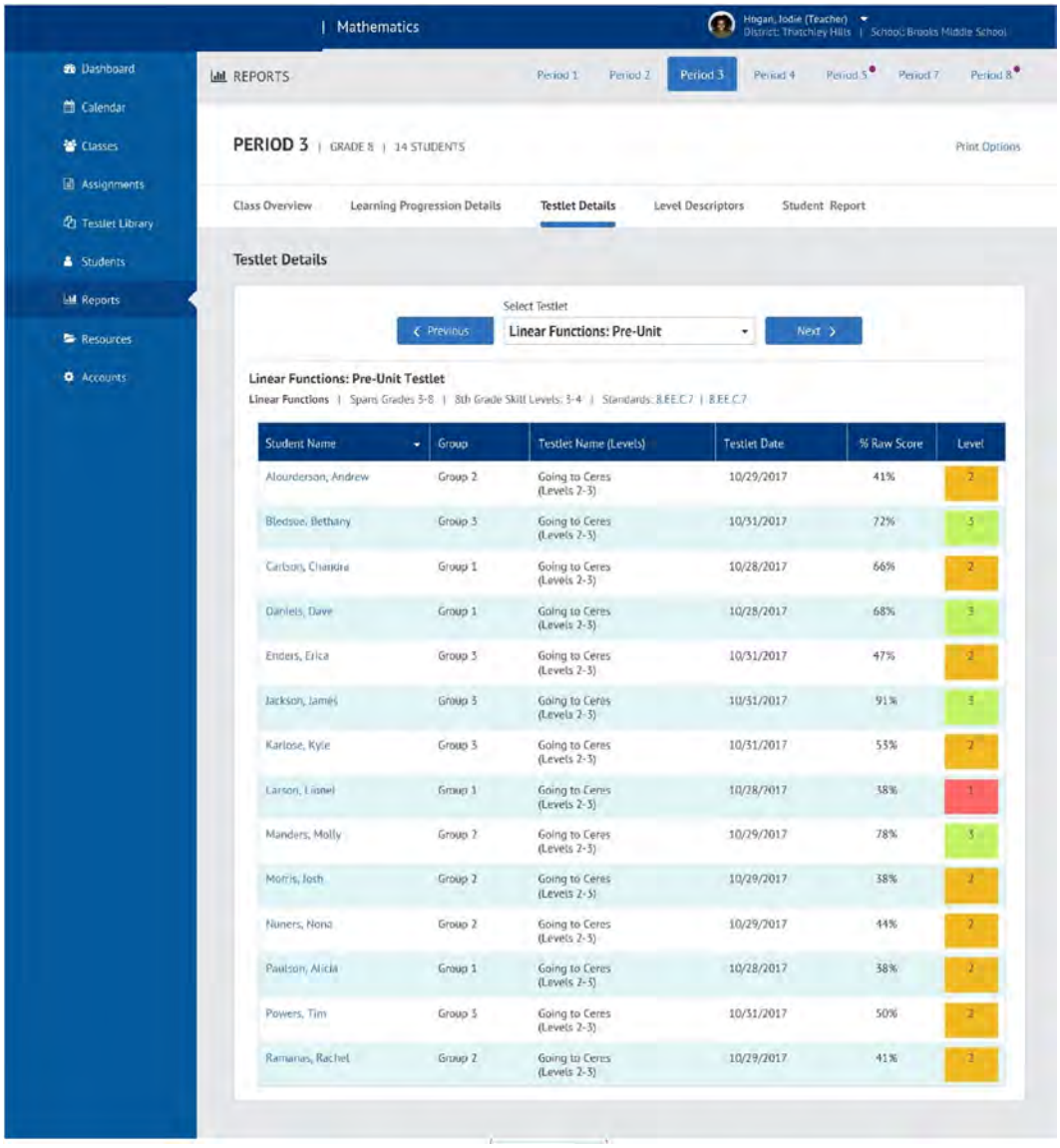

Figure A1 Class-level results are shown for one levels 2 – 3 mathematics interim assessment for grade 8 on the linear functions learning progression.

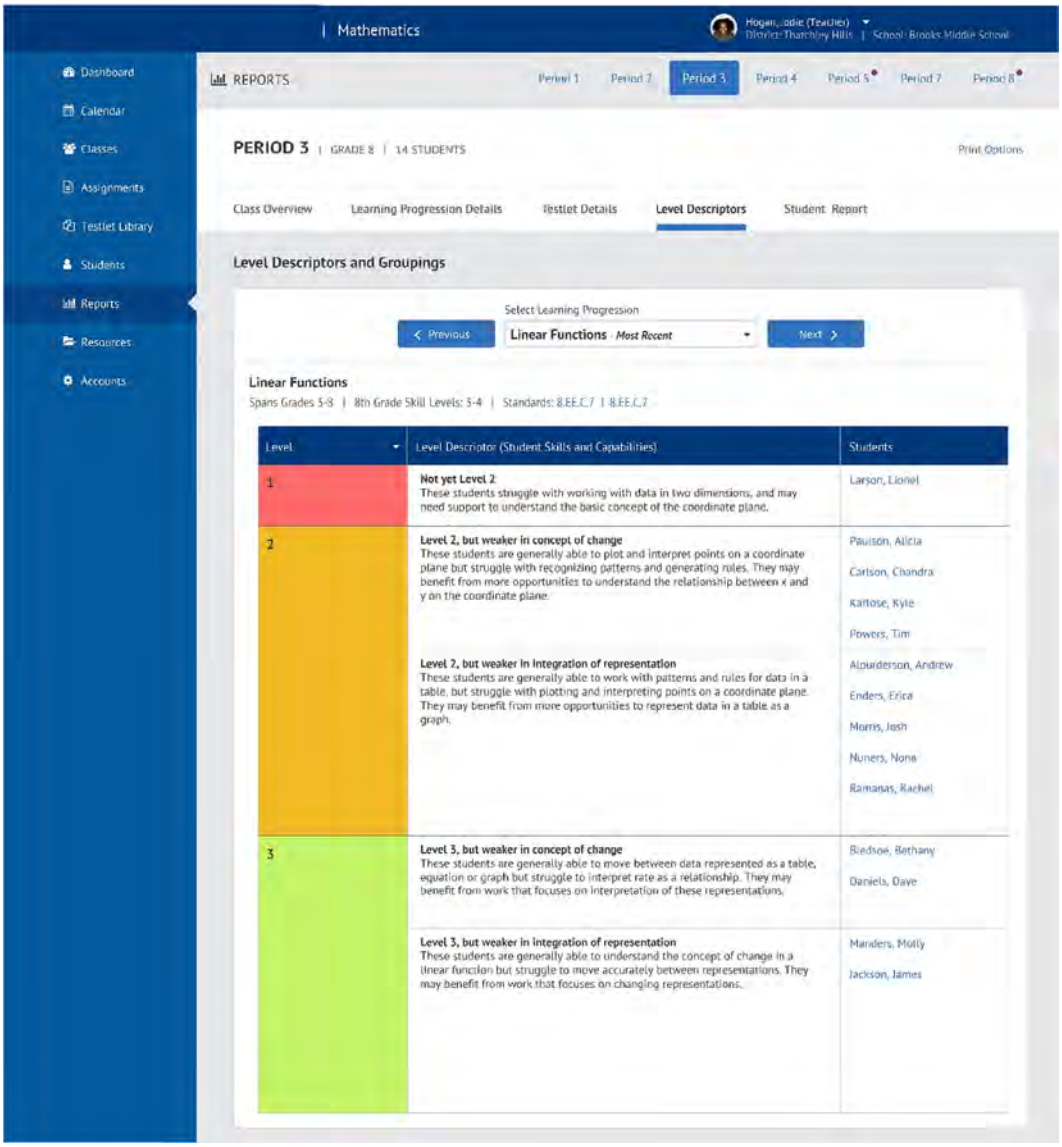

Figure A2 Detailed narrative feedback is shown for one levels 2–3 mathematics interim assessment for grade 8 on the linear functions learning progression. Students in the class are sorted into groups based on their LP level, which is further divided based on strength/weakness in one of two underlying progress variables.

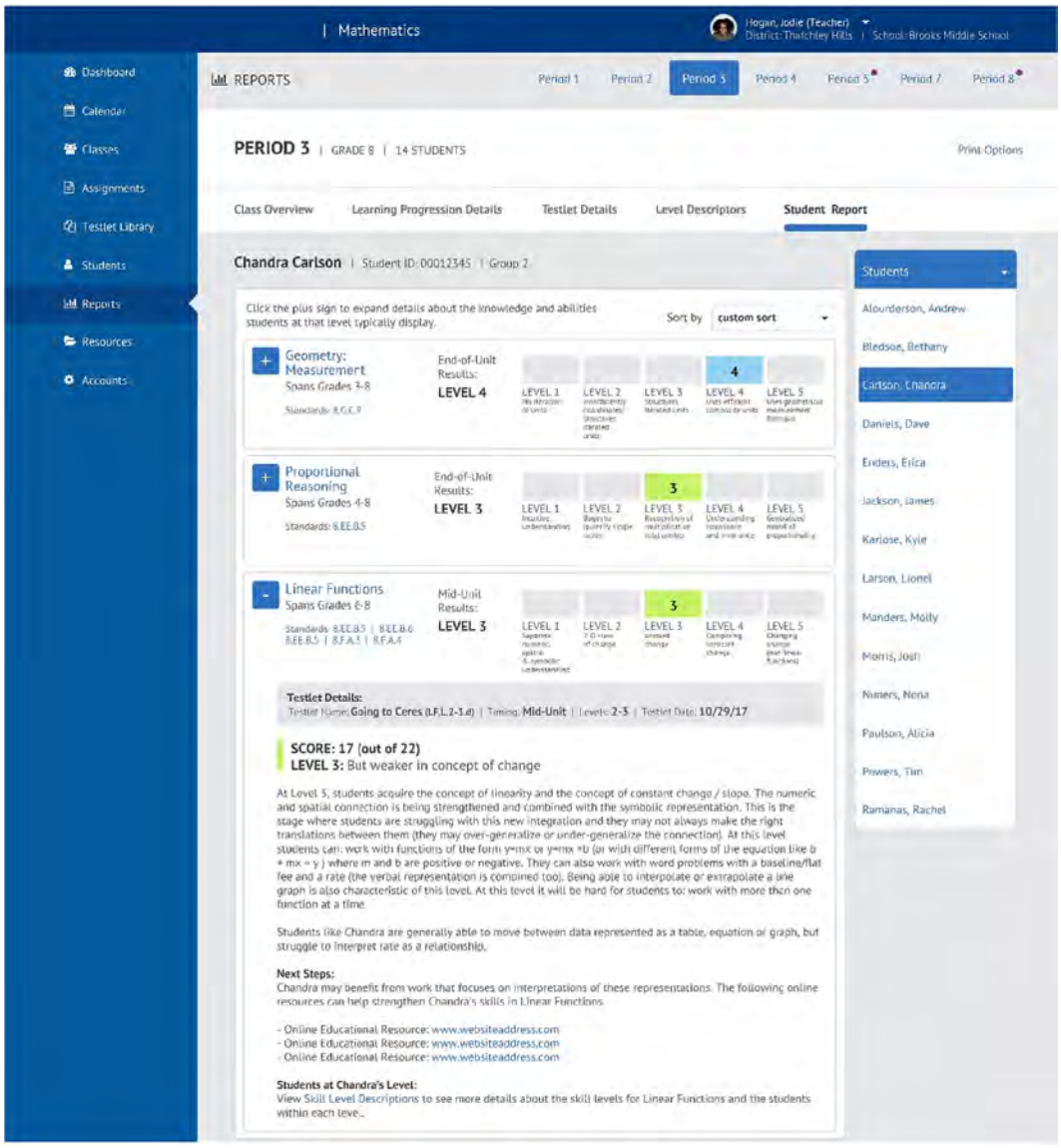

Figure A3 Detailed results are shown for one student in the class. In this view, the details for the linear functions learning progression has been expanded out. Detailed narrative feedback is provided for the student at her level, and targeted next steps and links to open educational resources that are targeted at the student's level are provided.

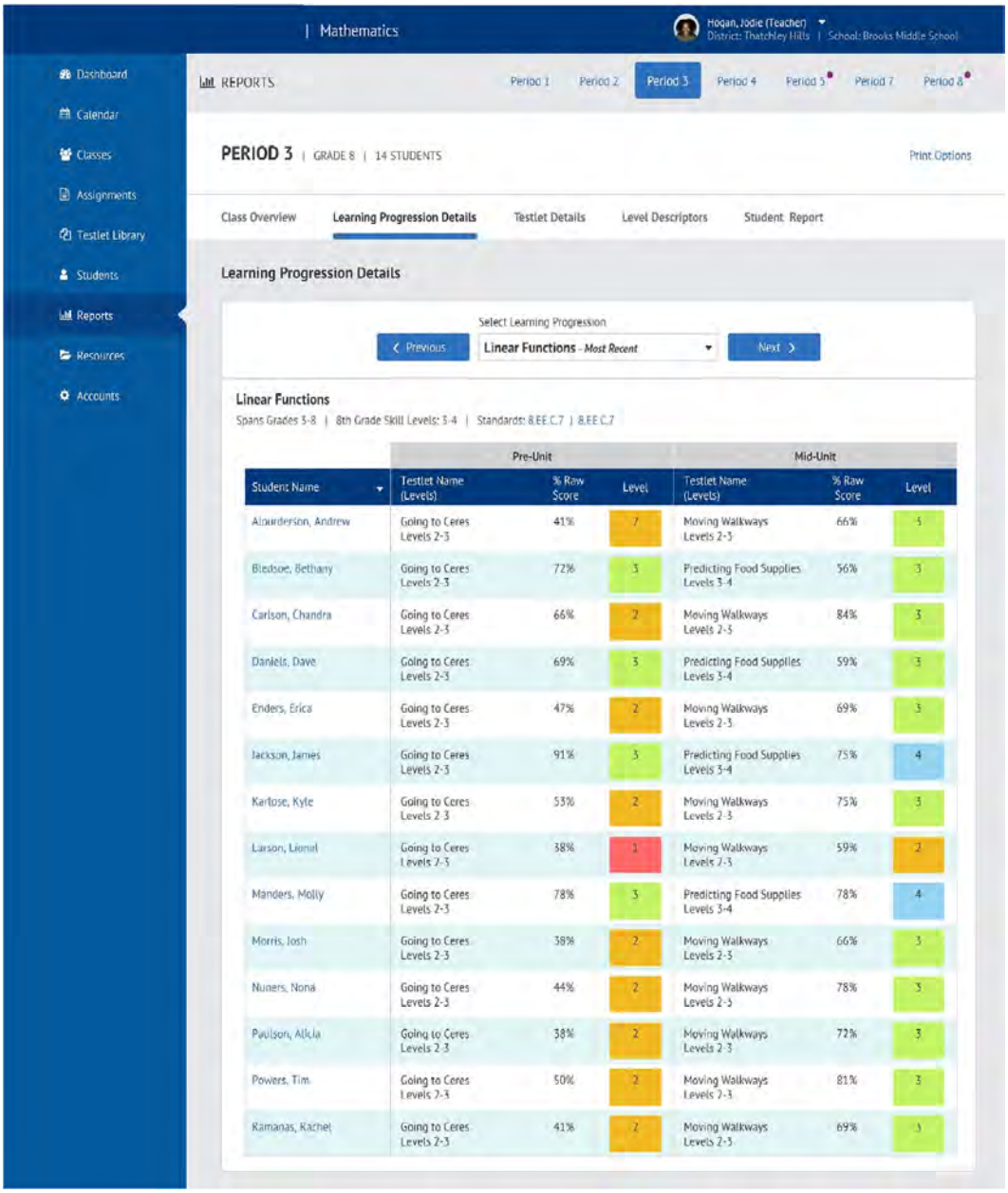

Figure A4 Class-level results for one learning progression are broken down into preunit and mid-unit results.

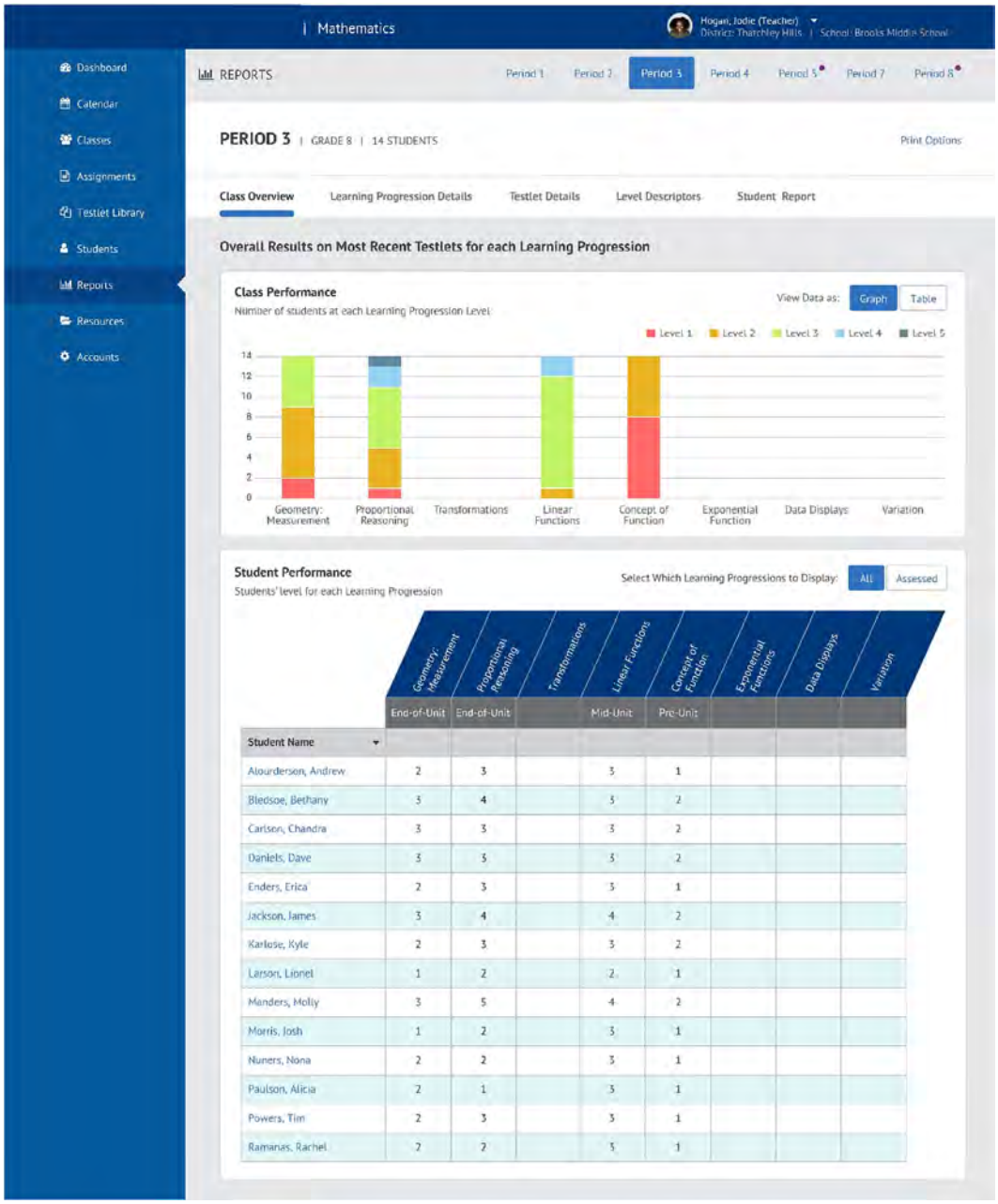

Figure A5 Class-level results are shown across all grade 8 mathematics learning progressions (LPs). The table shows the student roster information across the LPs for the most recent assessment (preunit/mid-unit/end of unit) that was administered for that LP.

English Language Arts Mock-Ups Designed for a Grade 8 Unit Assessment (Healthy Sleep: A Level 2–3 Unit Assessment on the Key Practice: Conducting Research and Inquiry)

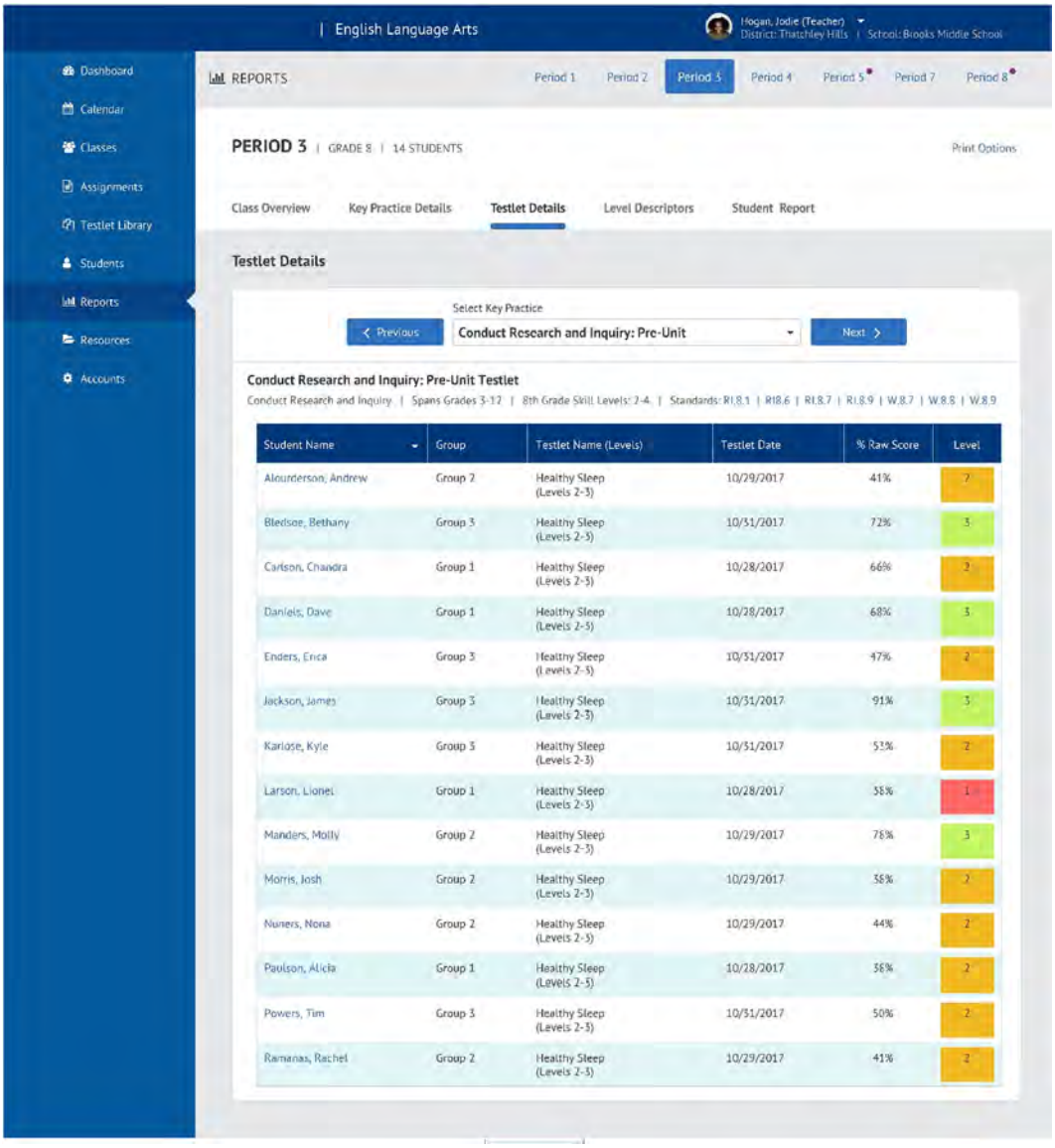

Figure A6 Class-level results are shown for one level 2–3 English language arts unit assessment for grade 8 on the conducting research and inquiry key practices.

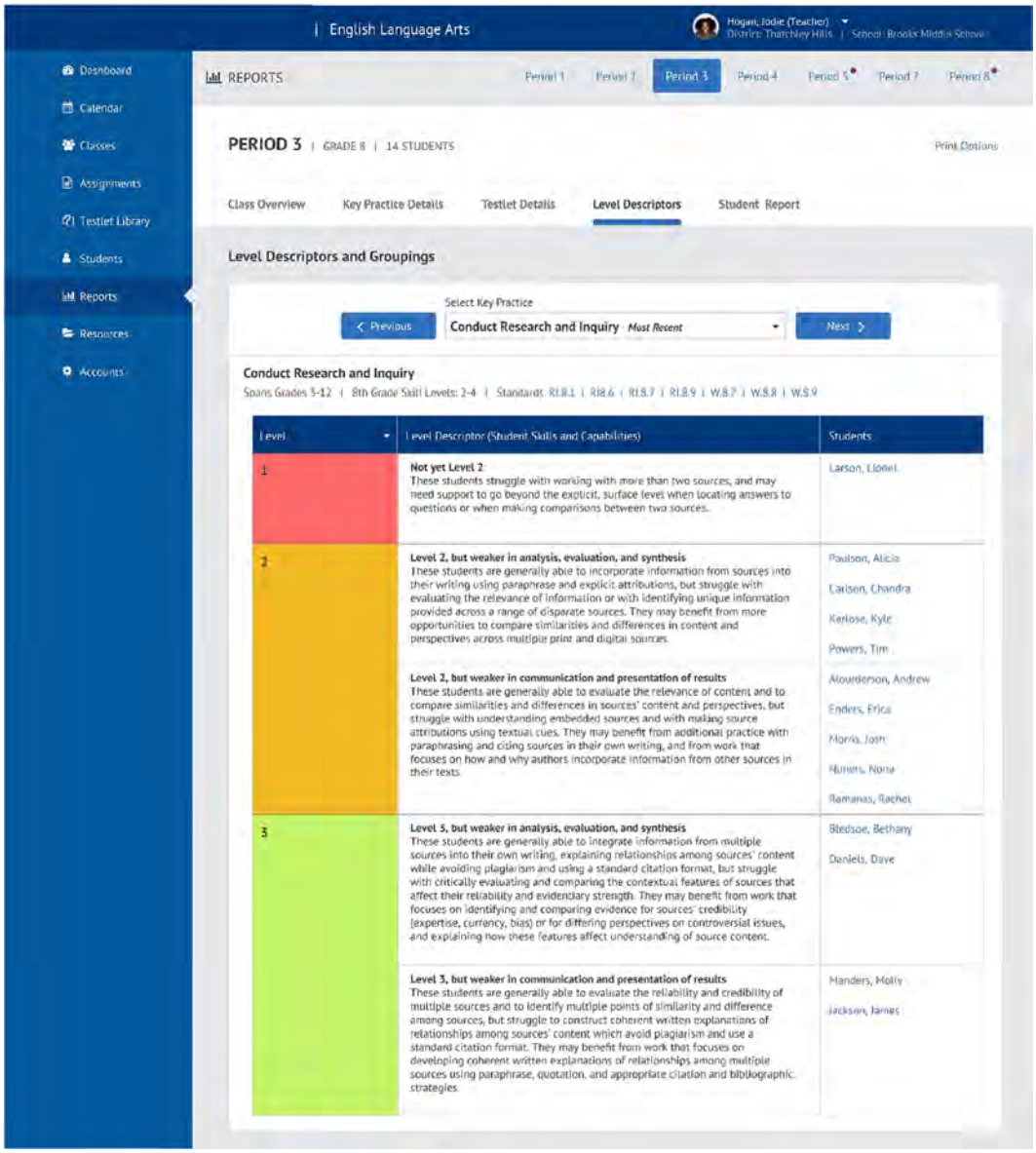

Figure A7 Detailed narrative feedback is shown for one level 2–3 English language arts unit assessment for grade 8 on the conducting research and inquiry key practice (KP). Students in the class are sorted into groups based on their level placement on the KP progression, which is further divided based on strength/weakness in the underlying phases of the KP.

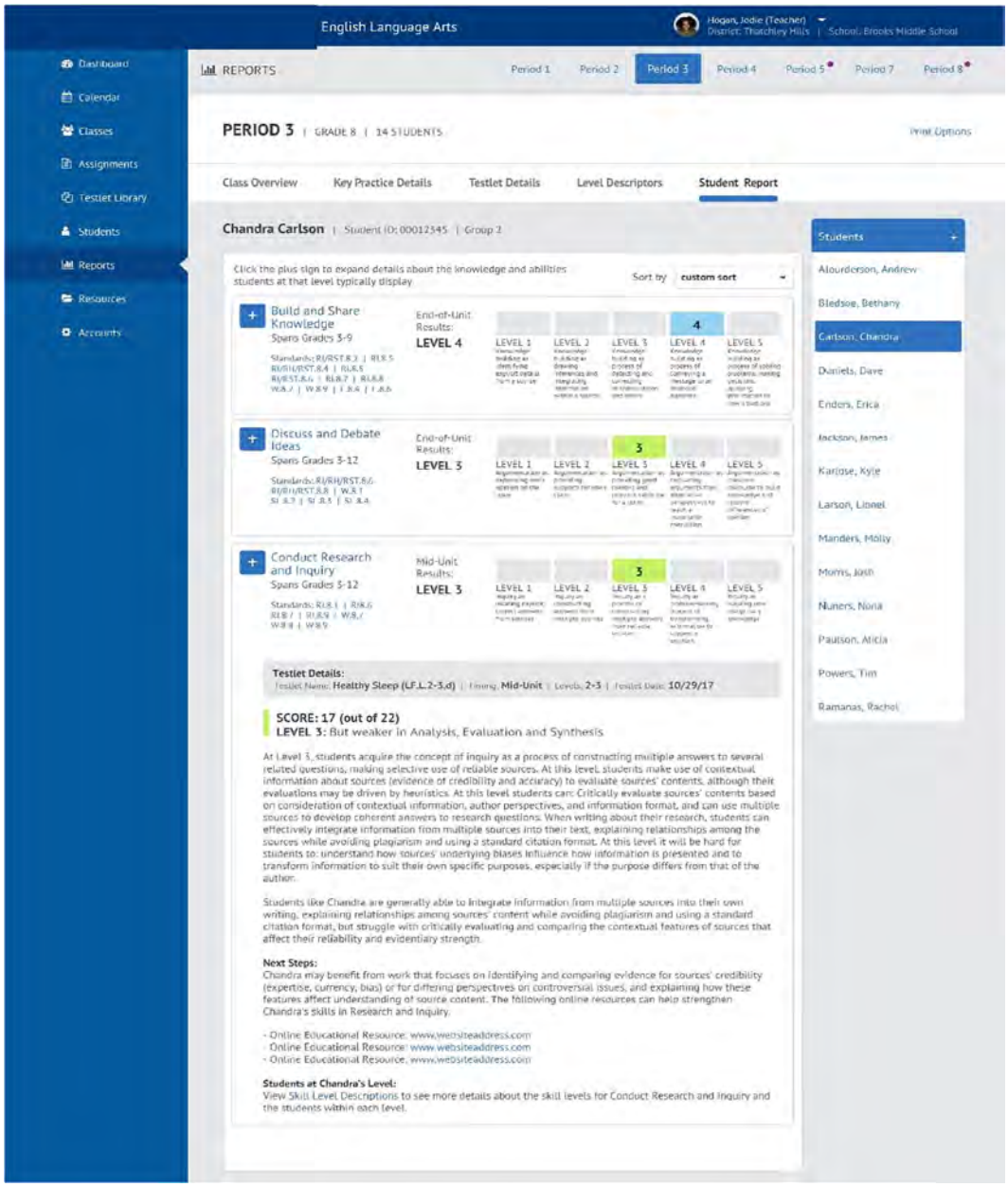

Figure A8 Detailed results are shown for one student in the class. In this view, the details for the conducting research and inquiry key practice have been expanded. Detailed narrative feedback is provided for the student at her level, and targeted next steps and links to open educational resources that are targeted at the student's level are provided.

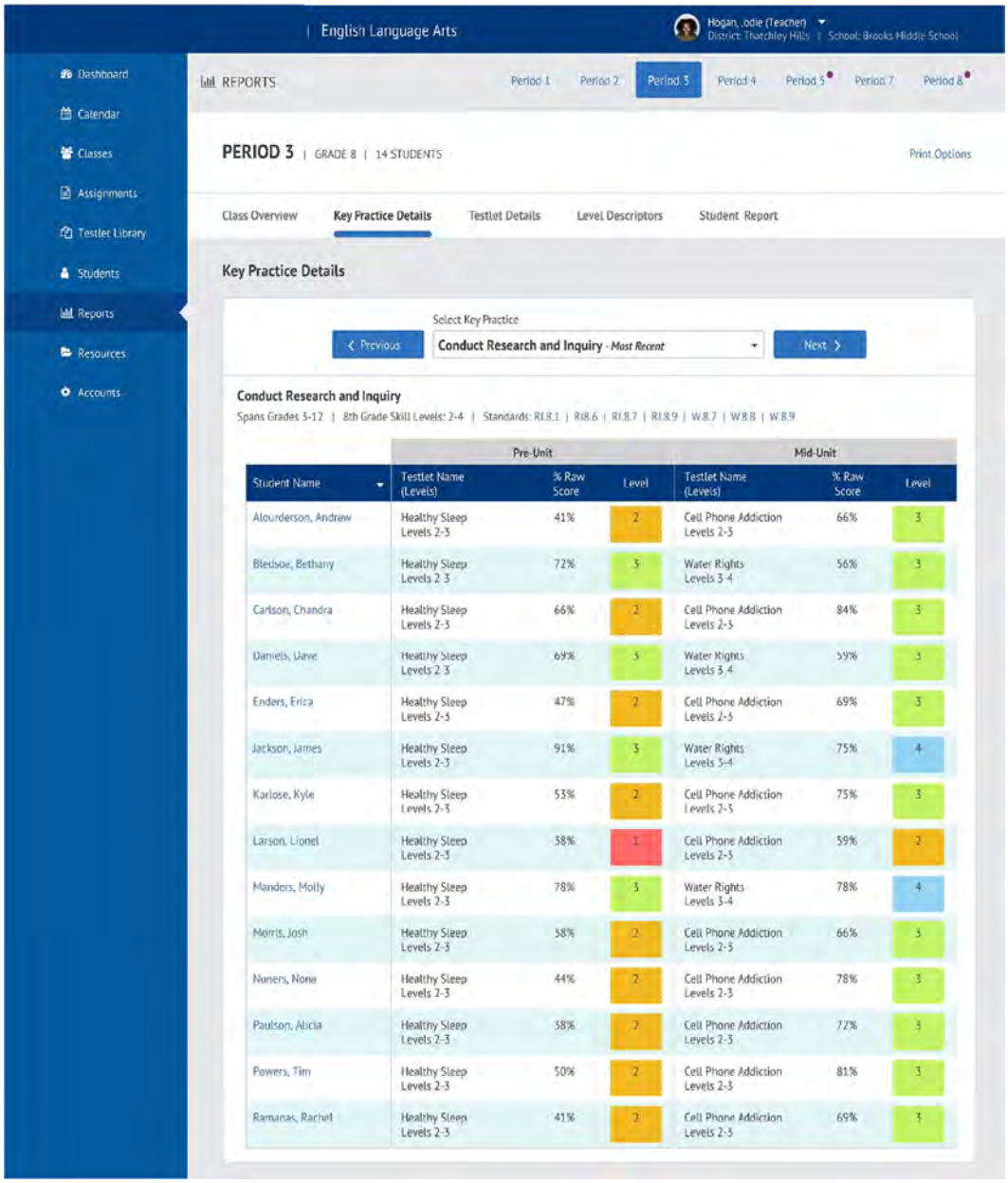

Figure A9 Class-level results for one key practice are broken down into preunit and mid-unit results.

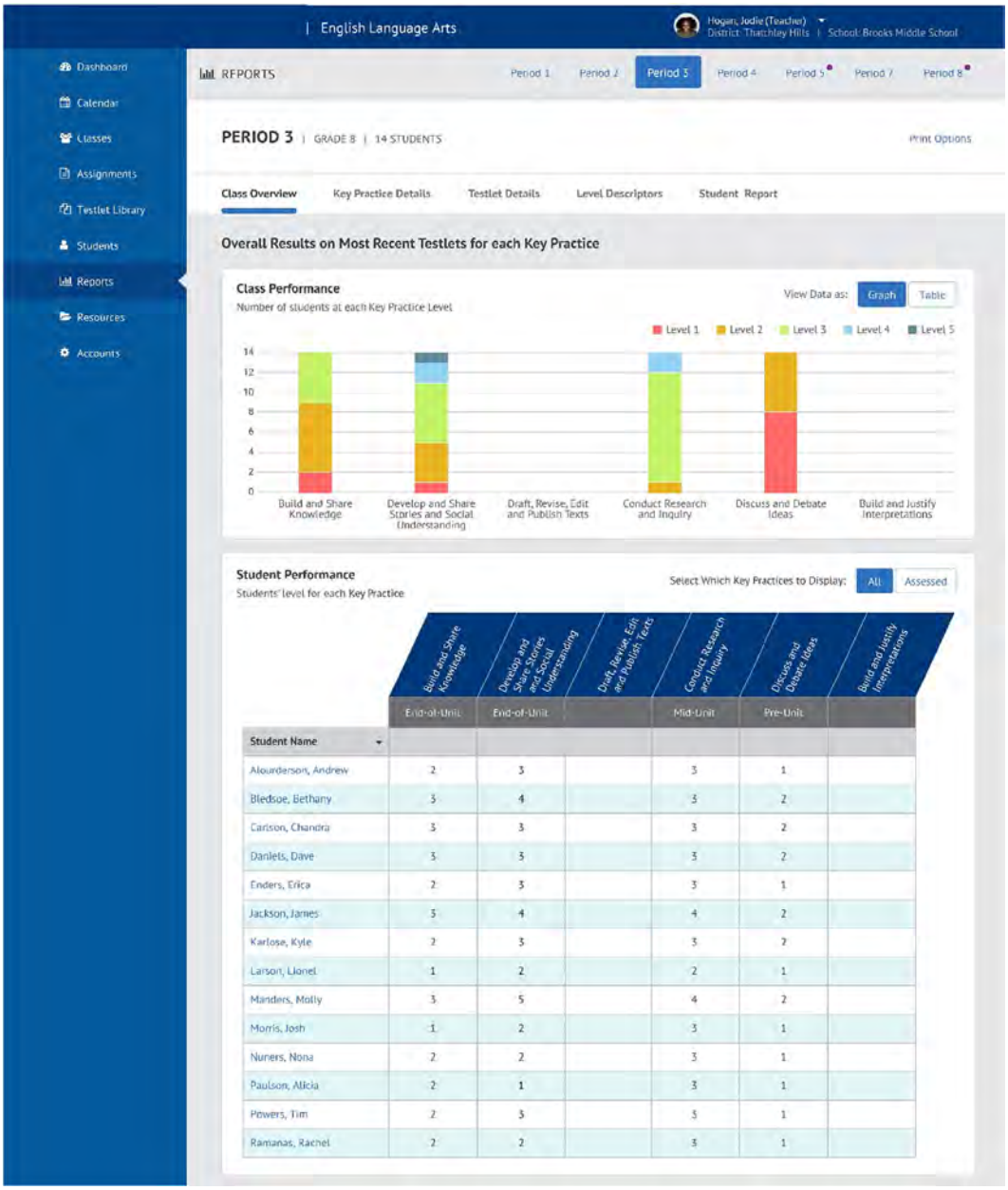

Figure A10 Class-level results are shown across all grade 8 English language arts key practices (KPs). The table shows the student roster information across the KPs for the most recent assessment (preunit/mid-unit/end of unit) that was administered for that KP.

Suggested citation:

Kannan, P., Bryant, A. D., Shao, S., & Wylie, E. C. (2021). *Identifying teachers' needs for results from interim unit assessments* (Research Report No. RR-21-08). ETS. https://doi.org/10.1002/ets2.12320

Action Editor: Heather Buzick

Reviewers: Malcolm Bauer and Diego Zapata-Rivera