CADRE REPORT

Teacher Reactions to the Phonics Content Referenced Growth Reporting Prototype: Findings from Interviews

By Nicolás Buchbinder, Olivia Cox, Erik Whitfield*

December 2023

Prepared by the Center for Assessment, Design, Research and Evaluation (CADRE) at the CU Boulder School of Education.

**Authors listed in alphabetical order, all contributed equally to the writing of this report.*

School of Education
UNIVERSITY OF COLORADO **BOULDER**

Introduction

In this qualitative study, we conducted a series of think-aloud interviews with in-service elementary reading teachers as they reacted to a prototype for reporting content-referenced growth (CRG) within a commercially available diagnostic assessment, Curriculum Associates' i-Ready Diagnostic. As in our prior work on content-referenced growth (Briggs et al., 2023), the design and structure of the reporting prototype is organized around a learning progression (LP). LPs afford rich opportunities for making sense of student growth by describing how a student's understanding of a 'big picture' concept in reading becomes more sophisticated over time with the right curricular and instructional support (Clements & Sarama, 2004; Lobato and Walters, 2017). Underlying any given LP is a theory-based hypothesis about how students develop knowledge and skill within the focal concept. However, adopting an LP approach to growth also requires adequate empirical support. In this report, we consulted an array of theories of word reading and reading development to define an LP about phonics and then validated the LP by exploring associations between its proposed levels and the difficulty of i-Ready Diagnostic items.

With its focus on phonics, the LP displayed in the prototype aims to describe the increasing sophistication with which children can connect letters in print to sounds in spoken words, as their knowledge of the relationship between letters and sounds becomes more comprehensive. The three levels of the LP—Partial, Complete, and Consolidated—thus communicate the types of connections between letters and sounds that predominate as children grow in their phonics knowledge. Our purpose in this study was to obtain feedback about the LP and its levels from practicing early elementary (K-3) teachers who are already familiar with the i-Ready Diagnostic. In doing so, we hoped to learn more

about three issues in particular: how useful it might be to embed LP information into i-Ready Diagnostic score reporting, how the prototype might support substantive inferences about student growth, and how usable the prototype appears to in-service teachers. These general areas of interest motivated the specific research questions in this study:

1. How can an LP for phonics facilitate meaningful connections between scale scores on the i-Ready Diagnostic and the types of letter-sound connections students have formed?
2. How are teachers able to understand and interpret the LP by using the prototype?
3. How do teachers envision using information provided by the prototype?

Methods

Participants in this study were recruited by Curriculum Associates. The process started with a short survey that was sent to a large group of educators and included questions regarding:

- Educator's school district geographical location
- Grades and subjects taught in 2023-2024
- Years of experience teaching ELA
- Years of experience using i-Ready Diagnostic
- Demographic information about the educator (gender, race/ethnicity)
- Demographic information about the students in their class (race/ethnicity, FRL eligibility, and English language learner status)

Ten educators that expressed interest in participating in the interviews were selected by Curriculum Associates because their background information satisfied some previously agreed criteria, such as representation of the K-3 grade range, and geographical and teaching experience diversity. Three of the educators were not able to participate at their scheduled time. Table 1 below contains information for all seven educators in the final sample. (All names shown in Table 1 are pseudonyms.)

Table 1. Information on Teachers in Sample

Name	State	Grade	Role
Alison	NJ	K	Classroom teacher
Kate	AL	2, 3	Classroom teacher
Amelia	GA	1	Classroom teacher
Deena	FL	3	Classroom teacher
Amy	FL	2	Classroom teacher
Bianca	CA	2	Classroom teacher
Kasey	CO	2	Classroom teacher

For this analysis, we conducted two separate rounds of coding to arrive at our findings across the seven interviews. The first round drew on a deductive approach, while the second involved inductive strategies for coding. Together, these two coding cycles allowed us to generate the themes outlined in the Results section below.

To deductively code, researchers define descriptive themes or categories prior to analysis that they expect to be relevant based on their research questions, and then identify excerpts of qualitative data that are indicative of those anticipated codes (Maxwell, 2012; Saldana, 2015). In the case of this study, we designed our teacher interview protocol with many of these deductive themes in mind, crafting questions that we hoped would spark teachers' reflections on how the prototype might facilitate meaningful interpretations about differences in i-Ready Diagnostic scale scores (RQ 1), as well as how understandable the prototype is to teachers (RQ 2) and how useful teachers perceive the prototype to be (RQ 3). Teachers' responses to those questions pointed to four deductive themes, which we describe in the next section with illustrative examples from interview transcripts. Our team identified these examples by having one researcher review transcripts and video recordings, annotating

the transcript for moments when an anticipated deductive theme arose, and finally transcribing quotes ranging from one to several sentences as evidence of interviewees' attitudes and perceptions on each theme. The research team then reviewed these findings together.

To complement the findings from the deductive coding, we also engaged in an inductive coding process (Kennedy and Thornberg, 2018), where we created categories based on the data produced in the interviews. This is a "bottom-up" approach in which researchers review the raw data (in our case, interview transcripts) to identify themes or categories that are common across instances. As opposed to deductive coding, the codes that emerge from an inductive coding approach are not necessarily aligned with the questions that were explicitly asked, allowing us to capture opinions and feelings of our interviewees that we were not necessarily expecting. Our process consisted of splitting the interviews across three independent coders who generated themes coming from each of the interviews using video recording and text transcripts. Each coder extracted quotes from the interviews that aligned with the codes they generated. Once each interview was coded, the research team came together to discuss the themes in each interview, find common aspects across different codes, and define a final set of codes. After that set of codes was finalized, each coder went back to the assigned interviews to reallocate quotes to different codes if needed and find quotes for the codes that were not originally generated from that interview. The final set of codes is presented in the Results section as three inductive themes.

Results

Deductive Coding

In deductively coding teacher interview transcripts, we identified four themes, each of which represents some idea, topic, or concept that we expected to surface when we posed a question or questions from the interview protocol (see Appendix A for the full protocol with questions and follow up prompts). The four themes include: (1) Connections between i-Ready scale score and LP levels, (2) Teachers' interpretations of differences between LP levels, (3) Prototype supporting inferences about

individual student growth and progress, and (4) Teachers' vision for using the prototype. Below, in Table 2, we note the number of interviews–up to seven–in which we found evidence related to each of these themes. We organize evidence types into categories of strongly and weakly positive or negative.

Positive evidence means that we heard a teacher strongly indicate some way in which the prototype would be supportive in relation to a given theme. Negative evidence are statements or comments from teachers that reflected room for improvement. Moreover, we interpreted statements as strong versus weak types of evidence depending on how quickly or directly a teacher made any interpretation related to a theme. If, for example, a teacher interacted with a feature of the prototype and immediately described the relationship between a certain scale score and students' level of phonics knowledge, we flagged it as strong evidence of the first theme. If, however, the teacher required prompting to make a similar inference, went in another direction, or left out elements of what we might consider a desired inference, we considered their response an instance of weak evidence. Table 2 outlines the frequency and strength and nature of evidence relative to each of the deductive four themes.

Table 2. *Count of Interviews in which each Deductive Theme was Identified*

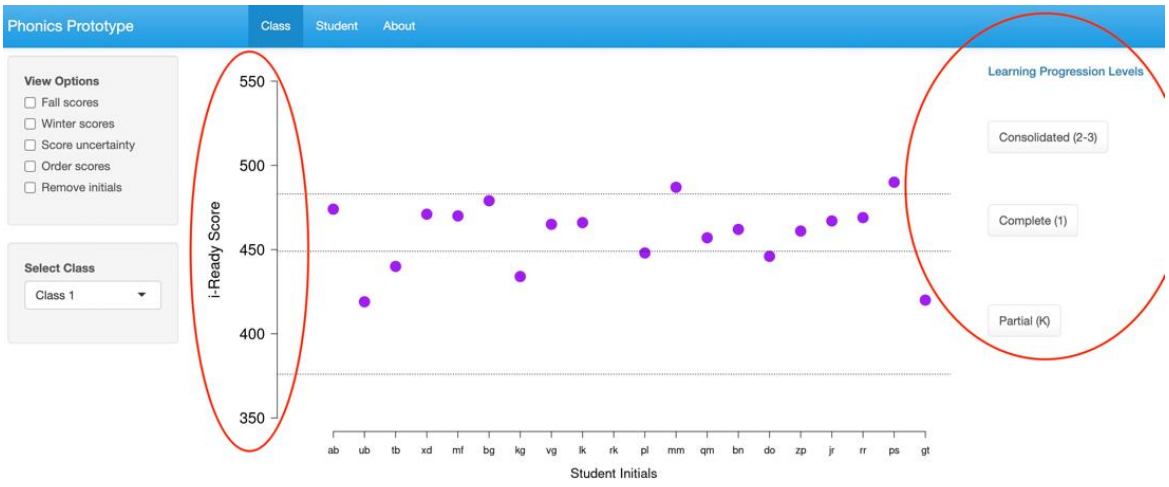
	Positive		Negative	
Deductive Theme	Strong	Weak	Strong	Weak
Connections between i-Ready Scale Scores and LP Levels	2	3	0	8
Teachers' interpretations of differences between LP levels	4	4	1	3

Prototype supporting inferences about individual student growth	3	3	0	2
Teachers' vision for using the prototype	3	1	0	2

Theme 1: Connections between iReady scale score and LP levels (Questions 1 and 2)

The first theme involves the connections teachers made between the i-Ready scale score, which appears on the left side of the y axis, ranging from 350 to 550, and the levels of the phonics LP, which appear on the right side of the y axis, labeled as Partial, Complete, and Consolidated (see Figure 1 below).

Figure 1. *iReady Scale Score and LP Levels in phonics prototype*



We expected teachers to connect the scale score range to the Partial, Complete, and Consolidated LP levels particularly when we asked them the first two questions of the interview, although comments related to this topic arose throughout the interview in many cases. To support teachers in thinking aloud about the relationship between the i-Ready scale score and LP levels, the interview facilitator sometimes provided a brief description of how and why our team decided to place the dotted horizontal

lines in the positions where they appear on the prototype (see a script for that description of the horizontal lines after Question 1b in the protocol in Appendix A). We found that when responding to Question 1 and 2 and reacting to what we shared about the horizontal lines, teachers tended to connect points on the scale score and LP levels in one of two ways, either by assuming that the connection between a certain scale score and a LP level also represented grade level proficiency, or where students "are supposed to be," and by reflecting on their own classrooms and experience to think through where students fall across LP levels and i-Ready Diagnostic scores.

In the first type of interpretation, teachers centered grade level expectations when connecting the i-Ready scale score on the left with the LP levels on the right. For example, when looking at the horizontal lines that delineate the upper and lower bounds of the Complete LP level, Amy remarked, "it looks like, maybe the two lines are the grade level expectation." In much the same way, Kasey's thoughts on what the horizontal lines symbolized were, "I'm guessing that band is that where the proficiency is, or they should be, like with the dotted lines. That's how I would read that, I guess." Kasey and Amy thus registered the existence of the lines that created rectangular areas or "bands" on the graph. However, rather than reflect on how those bands represented a level of students' phonics knowledge, these teachers' first instinct was to see the regions of the graph in terms of what they might communicate about on, below, and above grade level categories.

In a second variation on this theme, teachers drew on their experience or considered their current class of students to start to make sense of the relationship between scale scores ranges and the Partial, Complete, and Consolidated LP levels. Kasey, for instance, again referenced her expectation that the graph was communicating information about grade level expectations before reflecting on the level of phonics knowledge that likely predominated in her current class of second graders:

> I would think this class to me looks how many there are mostly proficient, or at least are in that band of where it's Complete. Yeah. And even these kids who are below aren't that far below. So,

> I mean, I'm thinking right now in my own classroom there might be some below this bottom line.

In response to Question 2 about how she would interpret the information presented about the class of hypothetical students included in the prototype, Amy also referenced her current group of second graders:

> Yeah, because at least in my experience with my second graders for them to be at Consolidated is... I maybe have three that are there now, and most of them are somewhere in between, Partial and Complete. And so, I think this class looks pretty good and that I would, you know, really focus on helping them learn how to orthographically map and like, be very intentional with, you know, why we do what we do and how our brain helps us see them [groups of letters] as parts to kind of get them thinking more along the lines of a student that would be considered Consolidated.

Unlike Amy and Kasey, who thought aloud about LP levels and their own classes, Bianca drew on another type of experience, her experience with the i-Ready Diagnostic itself and what students typically see and experience when taking the test. First, considering the scale scores with the same focus on grade level expectations that arose in other interviews, Bianca said,

> It feels like 450 is the average, probably, number for whatever grade that is because all students are like around the line 450 score... And then I can see that others have another, I guess, like another point 480, something like that, and that I'm assuming that's the highest level for those students.

After sharing her thoughts on this range of scale scores on the left side of the y axis, Bianca's eyes moved to the right side, with the clickable buttons for each LP level. Once her focus turned there, she had additional reactions, especially to the example items that appear when one clicks on an LP level button (see Figure 2).

Figure 2. *Class view upon clicking the Complete LP Level label*

Bianca began to make sense of LP levels in terms of how the example items included "audio and visual support." She noticed that there were pictures of the target word included on the example items for the Partial and Complete levels (e.g., the picture of a tail in Figure 2), but not for the Consolidated level and described how that kind of variation from level to level matched her experience with the Diagnostic. As she clicked on the LP level labels and perused the example items, Bianca explained:

> I'm assuming the test will give them images for the previous one [the Complete LP level], because that's what I see. And if they are Consolidated, they only have words like they're expected to read normally like we do in class, and they have like right now what I'm doing. They have, let's say, cat, and they'll have a [picture of a] cat. So now they have to think it's a cat, but they need to spell it, and they'll have the visual support. Versus the last one, Consolidated, I'm assuming the test wouldn't give them a lot of visuals. It's more to find out whether they can actually decode.

In her real-time reactions to the prototype, Bianca thus transitioned from interpreting the scale score range in terms of grade level proficiency expectations (as did many other teachers) to connecting the

scale score range to LP levels according to differences in students' visual and auditory experiences of items associated with each level.

Theme 2: Teachers' interpretations of differences between LP levels (Question 3)

The second theme from our deductive coding involves how teachers interpreted the qualitative distinctions between levels of the phonics LP using information provided in the prototype, including: the clickable LP level labels in the class view with a brief description and an example item (see Figure 2 in the previous section) and the About tab, which provides a bulleted list of phonics skills, understandings, and common errors associated with each level (see Figure 3), as well as additional example items and sample follow up activities (see Figure 4).

Figure 3. *About Tab, Phonics skills, understandings, and common errors*

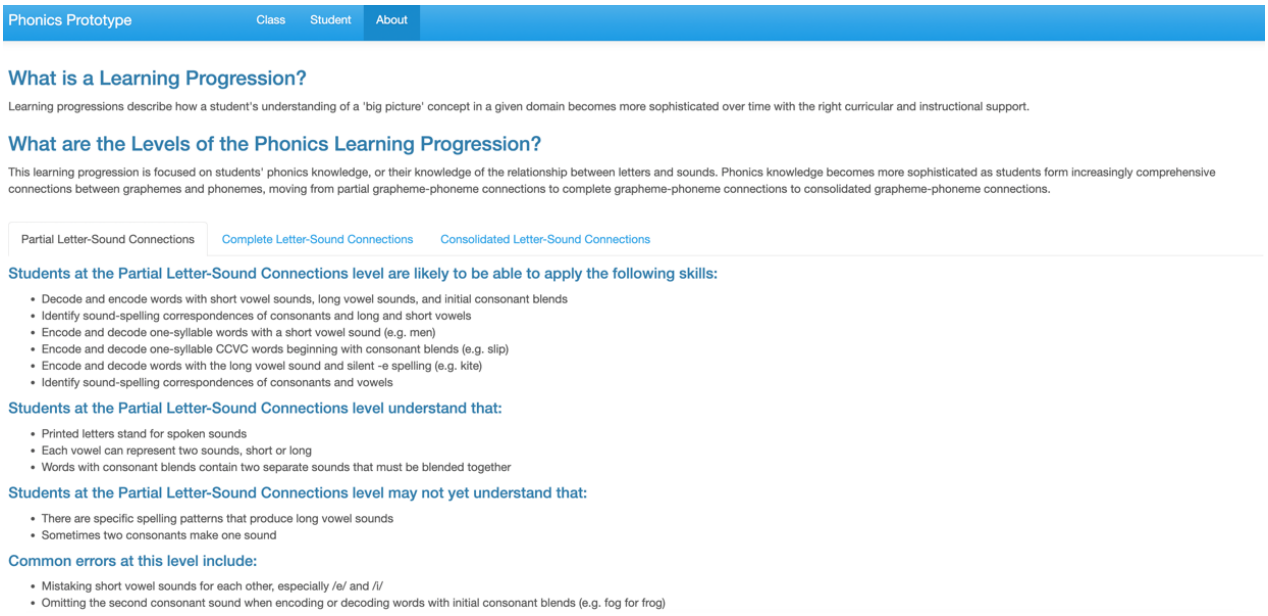
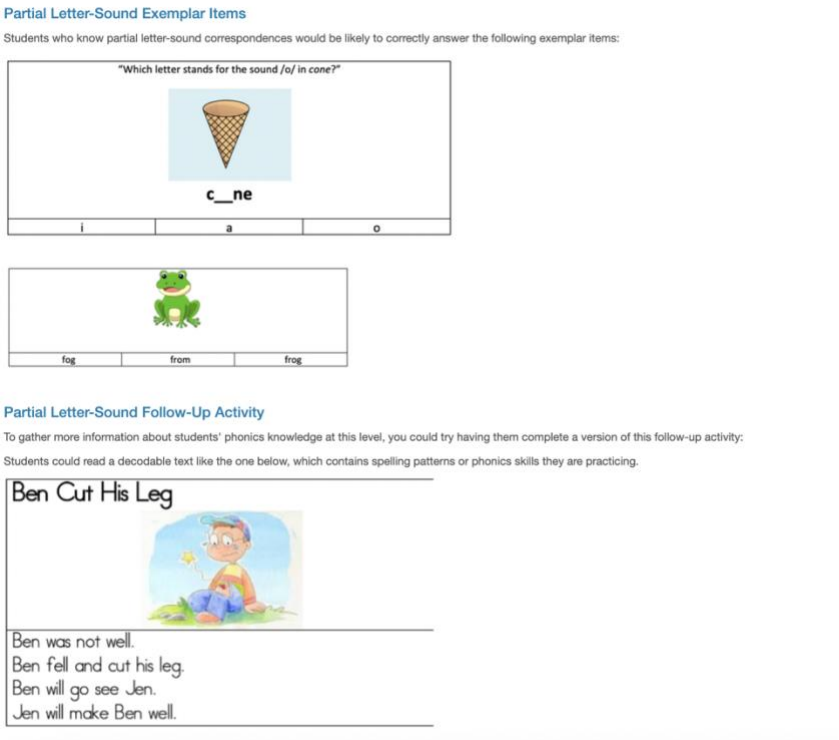


Figure 4. *About tab, Exemplar items and sample follow up activity*



Much like in the first two questions from Theme 1, teachers' responses around this theme tended to draw on their professional experience with current or past classes of students, multi-tiered systems of support, and reading curricula. Amelia, for instance, made sense of the LP levels by attaching them to different categories of intervention services students might receive: "Children at the Partial ones would be more of our Tier II friends, and then we would look at Consolidated or above is more in the Gifted." When we asked background questions in the first part of the interview, Amelia shared with us that she works in an intervention role, teaching students identified as needing Tier II support, or additional targeted instruction outside what is provided from core programming in the general classroom environment. Given her position and experience, it makes sense that Amelia would consider the LP levels in terms of the types of intervention services students receive.

When asked about what was changing across LP levels and whether it made sense to her, Kasey noticed how the bulleted list of skills, understandings, and common errors built on each other, with one

skill becoming more sophisticated over time. Using her mouse to circle the two bullet points for what students at the Partial LP level may not yet understand, she said, "It just makes sense that the Partial level they may not yet understand that and those pieces. But then that's exactly what showed up to be a Complete. That's what they need to be able to do." As an educator with close to 20 years of experience, Kasey's eye went to how the same core understandings change and become more refined from level to level. She later elaborated, "We've used so many reading curriculums over the years. But I do like how it [the bulleted list of skills on the About tab for the Consolidated level] talked about adding the endings and pieces like that because that's so hard for them. That's always a big piece." In reacting to the distinctions between LP levels, Kasey thus co-signed the inclusion of certain skills that she believed to be important, while also noticing the ways in which key phonics insights build across levels.

After reading the About tab, Amy's response to Question 3 drew on her experience with what she was "currently teaching," as well as her own professional knowledge about phonics and reading foundational skills. She explained,

> Your Partial letter-sound student is not as sophisticated of a reader or writer. They're definitely missing a lot. They may get beginning and initial sounds correct, but struggle with middle parts, or definitely, with longer words. Your Complete students are, gonna be more comfortable with longer words, with noticing the different syllables. And you know, like one of the things that I'm currently teaching is, you know, the six different syllable types. And so they're gonna be starting to be more familiar with that. Then, of course, the Consolidated is gonna be your student, that is really, you know, reading more fluently and just kind of like putting it all together without the labor of slowing down to decode.

This answer to Question 3 represents much of what we, as interviewers, ideally wanted to hear. We did not expect teachers to memorize the list of skills that characterized each level since those lists are quite long, but Amy captured much of what makes the Partial, Complete, and Consolidated levels qualitatively

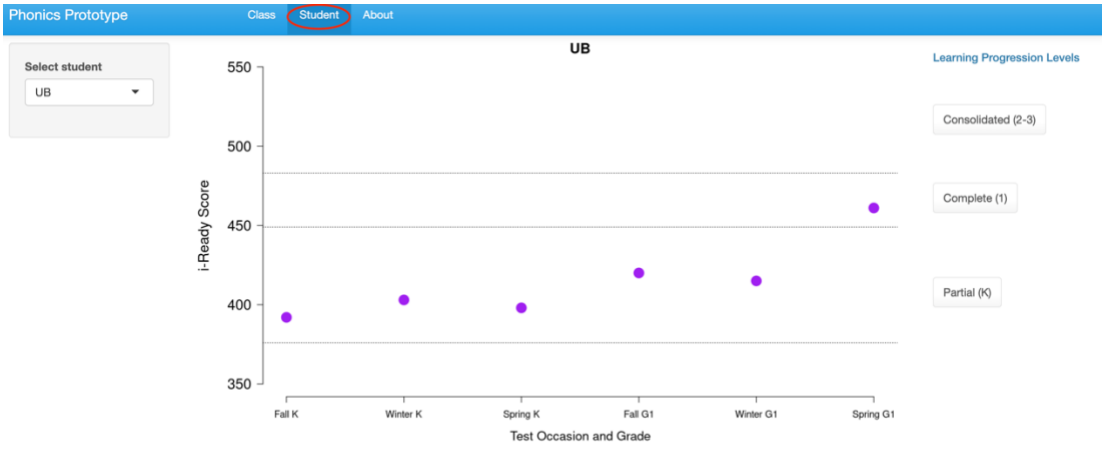
different from one another. In the introductory part of her interview, Amy shared that her school was piloting a new phonics curriculum and that she was immersing herself in the "Science of Reading," a "vast, interdisciplinary body of scientifically-based research about reading and issues related to reading and writing" (The Reading League, 2021, p. 6) . These two pieces of background may have contributed to Amy's response and the relative speed with which she picked up the main, high-level distinctions between levels of the phonics LP.

Amy, Kasey, and Amelia's comments all represented what we considered strong evidence of this theme about interpreting differences between LP levels, although they did so by drawing on different elements of their professional experience. As elucidated by Table 2 at the beginning of this section, there were also instances where the evidence of these interpretations of LP levels was weaker, with teachers seeming to not quite wrap their heads around what made Partial, Complete, and Consolidated letter-sound connections distinct from one another. For example, third grade teacher Kate never referenced any specific skill or idea associated with one of the three LP levels, instead offering more general comments like, "I know that them being able to associate the sounds with the particular letter is kind of a big deal right now" and "so, this looks like a normal phonics progression."

Theme 3: Prototype supporting inferences about individual student growth (Question 4)

The third theme involving teachers' inferences about individual student growth mostly arose during our conversation about Question 4 where we asked teachers to look at a sample student by clicking into the Student tab to display a student's i-Ready Diagnostic score on multiple testing occasions (see Figure 5 below).

Figure 5. *Student Tab in the Phonics Prototype for Sample Student UB*



Across interviews, teachers' descriptions of the sample student's growth from kindergarten to first grade tended to include two things: recognition of a "lag" in growth and discussion of a point in time where everything "clicks in" for students. Regarding the first point, Amelia and Amy both spoke about "summer slide," with Amelia explaining, "it seems like we've got some growth, definitely got that little lag from summer and then back growing again." Bianca noticed another pattern: "This one looks like the winter Kinder is higher than the spring Kinder. Same for fall first grade and winter first grade. Like the student, is kind of, you know, up and down." Bianca was one of the only teachers to refer to the phonics LP levels when describing student growth, saying,

> Yeah, it will be just going back to the partial, to the progression levels. Because I could see for the fall, winter and spring K, I would go to Partial to see all of the descriptions that they have there to see what they know or what they're missing. Same thing, now that I know that one [the Complete level] means grade one, I would go to fall, winter and spring for Complete to see What they know it from based on that Grade 1 level.

Thus, many teachers narrated what they saw in terms of "lags" in student UB's i-Ready Diagnostic results, and fewer teachers, like Bianca, discussed growth also in terms of the types of letter-sound connections UB likely knew or had learned at different testing occasions.

Teachers also identified a point when UB's growth started to take off, referencing how their own experiences with children learning phonics often involved a turning point where everything "clicks in" and accelerates growth. Looking at the difference between winter and spring of first grade, Deena remarked, "I'm looking at student UB. Low to start with, but then made a really big jump from winter to spring. So maybe that light bulb did click for that student." Kasey echoed this sentiment, saying, "sometimes it just clicks with kids... and you can see where it maybe clicked with this guy. Definitely a little flat line in kindergarten, I would say, and then, first grade… like I said, sometimes it just kicks in." Interestingly, Kate's comments mirrored Deena's and Kasey's, but then seemed to return to some of the concern with grade level proficiency expectations that predominated in teachers' responses to Questions 1 and 2. She said,

> It looks like their kindergarten [year], they made slight progress and then in the fall of first grade, looks like everything started coming together and they were able to make a little more or a little, I would say, significant progress. Winter, I would attribute that maybe to the holidays, and the days they have out of school. And then Spring, they jumped to almost grade level but not quite. So, it looks like, based on the data for this particular student, it took about, maybe a little over a year to get near grade level or on grade level.

Unlike Bianca, who was thinking aloud about where the student fell in terms of LP levels across different grades and testing occasions, Kate was tuned into how the student's growth in Diagnostic scores got closer and closer to grade level proficiency. In this way, the focus on grade level expectations thread throughout the interview, figuring into not just responses to the first two questions but manifesting throughout our conversations with teachers.

Theme 4: Teachers' vision for using the prototype (Questions 5-8)

The fourth deductive theme reflects all the ways in which teachers envisioned using the prototype, often sparked by questions from the interview protocol like, "Where, if at all, does this kind

of information fit within your instructional planning? What decisions would it help inform?" A number of teachers pointed to the role the prototype might play in constructing small groups and planning small group instruction. Amy explicitly said, "I would probably use it more as a small group kind of tool..., once I identify those students that I can clump together." Amelia similarly discussed how she might draw on information from different tabs of the prototype in preparing for small group reading instruction:

> Definitely for small groups, I would use like the Class tab, and just kind of look and see like I got UB way down here, and so is GT, like, I might want to put them together. And that's when I would probably jump back to the Student tab and say, Okay, well, where did UB score exactly?

Concurring with this sentiment, Deena said, "I could see it helping form small groups. It gives you a better picture of where they are for learning their phonics." She even pointed out that information from the About tab "could maybe give me differentiated homework for this kid, or some different morning work." Relatedly, Kasey remarked that the prototype could support teachers in "skills-based grouping," or dividing their class into small groups based on students' particular skills profiles rather than, for example, independent reading levels. Taken together, the teachers' comments suggested that the prototype may be able to make some contributions to instructional decision-making especially when it comes to teaching reading in small groups.

Inductive Coding

Our analyses produced three inductive themes: (1) Feedback on LP levels, (2) Making sense of graphs, and (3) Use of prototype for professional development. Each of these themes holds different types of feedback that teachers provided to the current version of the prototype, to which we return in the Discussion and Future Directions section, where we overview potential changes to the prototype.

Theme 1: Feedback on LP levels

In the interviews, teachers provided valuable (implicit and explicit) feedback about the LP that the prototype introduced. As discussed in the previous section on the first deductive theme, something that was clear in all the interviews is that for the educators the most natural interpretation of the information that we presented to them was that in some way the prototype was providing information on whether students were "at grade level". Rather than focusing on the scale scores or inquiring about the LP levels that appeared on the right of the class view, interviewees often jumped directly to interpret the bands that were marked in the scatterplot in terms of whether the students were attaining grade-level expectations.

As discussed previously, there was no explicit link in the protocol (at least in the version that was showed to the first five teachers) between the LP levels and the expectations for each grade in terms of learning in phonics. However, the educators' first approach to the information presented in the protocol was linked to the grade-level expectations no matter which grade the educator was teaching this year. We discussed two possible explanations for this behavior. First, teachers in these grades may have less specific disciplinary training in literacy and may prefer information that is more closely linked to the outcomes that they are expected to attain in children. This provides an interesting route for future exercises like the one we are conducting, as it is indicative of how much educators read the information that is presented in terms of the skills that they are responsible for, sometimes ignoring other information that is available for them in the same data visualization. Second, teachers might have been exposed to other assessment reports (including regular i-Ready reporting) that uses these categories, and they may be extrapolating these categories to our prototype. As one teacher put it,

> I was just thinking that because, like, I said, like, my brain works with like grade level below and above. So is just thinking, if I'm in second grade and I see these, can I really distinguish like whatever I'm teaching in second grade? Because it's like a lot of things. And I'm like, should I try

to work on all of these? Because then I could put my student in the Consolidated or you know, like, how would I know which is my span content?

One lesson from this finding may be that previous experiences in data interpretation structure how educators interpret new information that is presented to them. Taking this into account might be key to achieving success in presenting new information to educators.

Related to these interpretations, teachers mentioned frequently that information of a smaller grain size would be more useful to them, especially when they considered using the prototype to support small group instruction. Even after teachers had reviewed the learning progression level descriptors, they felt they needed more specific information to be able to make better use of the prototype for instruction. Amelia, for example, shared the following thought when reviewing the Student tab: “I assume there would be like broken down areas, like, whether it's phonics or phonological awareness, or what, or what part of phonics we're not doing well in.” Bianca was interested in knowing more about specific student response patterns on the i-Ready Diagnostic. As she put it,

> But I'm thinking, if there is a like a more extended report of what they missed, even like the questions, to see if there is a pattern of like… Oh, this student is missing the short i, or contractions, or something that it's a little bit more helpful. Because having the graph, yes, I see they're progressing. But having more specific explanations of where exactly do I need to look at? Because I know it. Usually, it's an overall from all of these questions. This is what I got as a student. But from a teacher point of view… It's like, yes, but what else can I do now?

It is possible that for some of these ‘big picture’ concepts, the grain size of the levels of the learning progressions that we developed as a basis for the prototype does not align well with the type of decisions that teachers have to make daily. In other words, although size and scope of the phonics LP levels are reflective of theory-based developmental stages through which children progress as they learn about an alphabetic system, the contours and boundaries of the levels may not be as actionable for

teachers as we imagined. From these interviews, it seems that teachers would at least appreciate having more information to understand progress within each level, and not only between levels.

Another type of feedback that we received was that the names, or labels that were chosen for each of the levels in the learning progression were not immediately recognizable. Amelia told us that,

> I feel like there might be like a better term that could be called. I don't know. Just, just looking at partial, complete, and consolidated. Just doesn't like, ring an instant bell for me, as to what I'm looking at.

It is possible that the interpretability of the prototype can be improved by using labels for the LP levels that are more readily interpretable for teachers. However, since partial, complete, and consolidated are all terms associated with an established developmental theory of word reading (Ehri, 2005), there is a risk that using new labels that may be more connected to daily teaching practice might also imply a departure from the terms and concepts that were originally coined by researchers.

In an additional piece of valuable feedback, teachers who had experience working in K or Grade 1 highlighted that the LP as it was showcased in the prototype did not capture all of their students. In particular, students that are below the Partial level – common in K and possible in Grade 1 – could not be classified in terms of this LP. As Amelia put it,

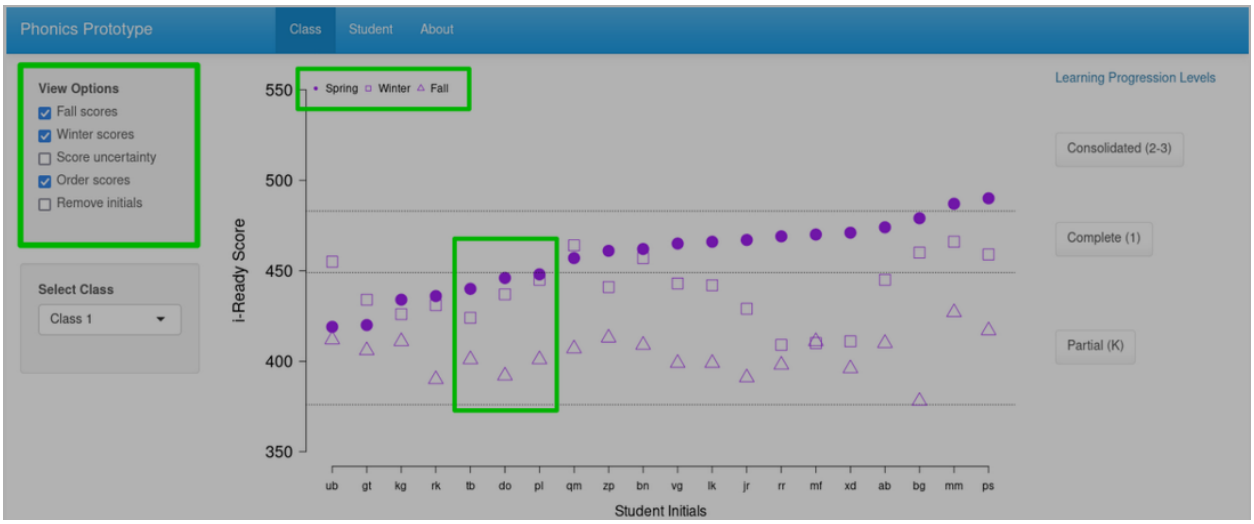
> I just feel like there might be something before partial, if that makes sense like a pre-reading, or I don't know, I don't know it would be that. But cause partial felt like it had a lot already like, we knew a lot already when we were just at, partial, so I'm just wondering how that would fit in for K... kindergarten learners, just... it just feels like maybe there could be another description below partial that would be, I don't know, emerging skills or something.

Comments like these, in addition to teachers' reactions to the LP level labels, prompted us to consider some revisions and updates to the prototype, which we describe in the Discussion section.

Theme 2: Making sense of graphs and suggestions for improving the prototype's interpretability

One aspect of the prototype that received significant attention from nearly all of the interviewees was the way that students' Spring, Winter, and Fall scores were represented on the plot in the class view. The teachers' comments tended to focus on the extent to which the plot points were easily distinguishable and the lack of interactive functionality (i.e., the fact that one cannot click any of the shapes below in Figure 6 to pull up individual student results on the Diagnostic).

Figure 6. *View options in the prototype*



With regard to how well the plot points allowed the teachers to distinguish easily between different students and different test occasions in the whole-class view, the feedback was mostly negative. One teacher, Alison commented at length on how difficult they found it to visually match the points in the plot with the student initials. They stated that

> This is definitely confusing because . . . I think it'd be hard to track up from each kid's initials at the bottom to see what their dots were, and if I was designing this I would do different colors for each kid. So you could see, like, or even a pattern like a rainbow . . . because otherwise, like, the dots could get confusing, like, you know C2, the first one, could be red and then UV could be blue, or, you know, just make it more clear.

This feedback explicitly indicates that the use of different colored points would make it easier for this teacher to match points to students. Given that they noted the difficulty of visually tracing from a student's initials to the corresponding point, the teacher seems to think that introducing different colors would simplify the task of matching different elements of the graph. Though, at least one teacher did not seem to have difficulty in this regard. Kate, upon first examining the whole-class view, said, "I would possibly assume that down here, where you have your student initials, these initials go with one of these purple dots at the top and will tell you where that particular student lies."

While Alison was referring to the difficulty involved in matching points to students, other teachers commented on how challenging they found it to make sense of the Spring, Winter, and Fall scores when shown on the plot simultaneously. For instance, Deena remarked that

> You've got everything with purple. Maybe put a different color, you know, for each thing. I mean, I can see the different shapes and all that, but maybe a different color . . . if there's a different color, you would see the different levels, maybe a little bit better . . . You know, they have a square and a triangle in the same spot, but if it was a different color, it'd probably be easier to see that.

This teacher clearly thinks that colors are easier to distinguish than shapes, and that it would also be easier to interpret overlapping colors. At one point, Kasey seemed to take a different stance when she stated, "So, what are the dots? Is that the spring one then? Gotcha. Okay, so I can see all their winter. Oh, see, I like that . . . I do like that, how it's laid out, because then you can see their growth" In contrast to other teachers' comments, Kasey found the presentation of different test occasions easy to understand, at least initially. Later in the same interview, this same teacher also said, "I almost wonder like if you're showing the like Spring, Winter, Fall, maybe different colors . . . if we're going for a visual, sometimes like colors help with that too." This statement is much closer to those of the teachers mentioned so far. Across all their comments, teachers tended to indicate that colors are easier to

interpret and distinguish than shapes. We discuss how this topic has motivated potential changes to the prototype in the Discussion section.

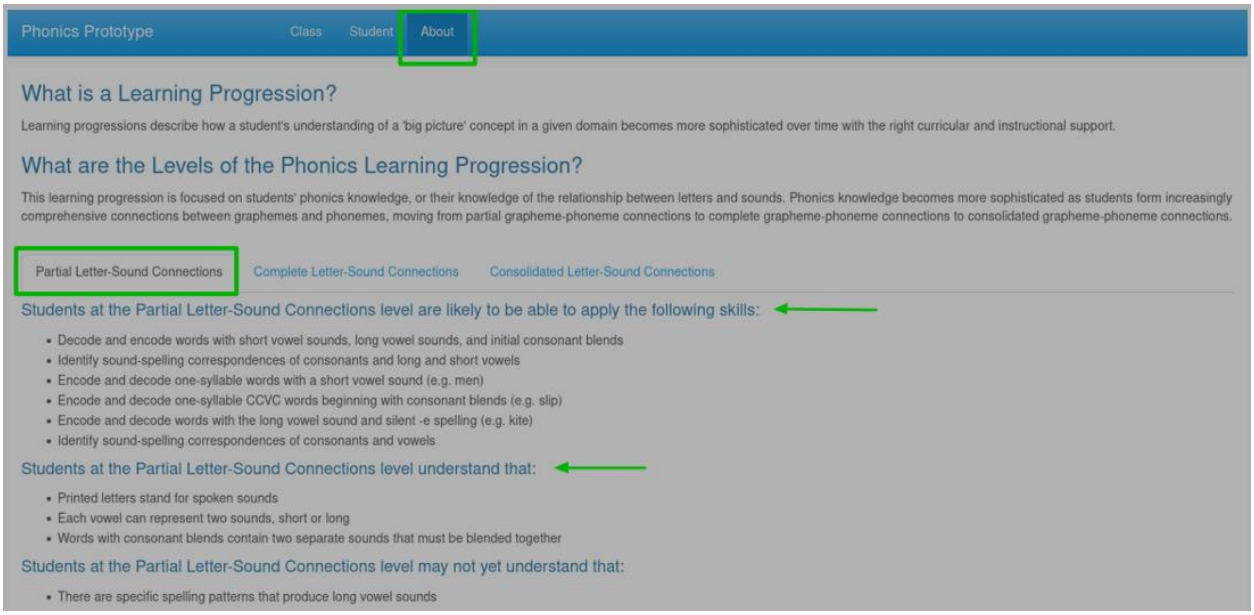
A second point related to the presentation of student scores in the whole-class view plot has to do with the lack of interactive functionality. When prompted to identify a function of the prototype that would display additional information about a particular student, Bianca said, "I'm assuming [the purple dot] would send me, which I guess it doesn't." Subsequently, they explained that being able to click on a point to access student scores "would be helpful because you know like I wouldn't have to . . . scroll down to find each student. I would just click on that one student and go straight to their report." Alison made similar comments, but instead of wanting the points of the plot to be interactive, they wanted the student initials to be clickable. Specifically, they said, "I wonder if I clicked on the initials if I get to see their report or not. But, I don't think the initials are clickable, so I can't do that. So, I guess I would have to go up to student to figure that out." They returned to this idea later in the interview stating that they thought "being able to click on the initials at the bottom would be really helpful so I don't have to like go up to "Student" and then click down to the student. I think that would make it more user friendly." Adding to the case for quicker access to individual student scores via the whole-class view, Amelia said, "I would just like to see the actual number . . . Just so I could see like the actual amount of growth . . . maybe if that was like a hover function or something." Taken together, the feedback provided by these three teachers provides convincing evidence that incorporating more interactive functionality into the plot would enhance the user experience and make it easier for teachers to access relevant information.

Teacher comments about color were not restricted to the display of individual student scores for different test occasions. There were also remarks about the potential for color to help teachers distinguish between the different levels of the learning progression itself. For example, when prompted to describe elements of the prototype that they particularly liked or disliked, **[Bianca]** said, "If I had like a color-coded bar . . . starting from the bottom line to the middle line, that color saying partial, I would

match it. So let's say it's light pink, and Partial, that little box is light pink. It would make more sense that all that range is Partial." Alison voiced a similar interest in color when stating, "I know in like the current i-Ready model they have like the shades of green, I think, for the different sections. So, even if it was like color-shaded, like the partial was a color, the complete with a color, the color consolidated, like all this whole area was shaded in." Based on these comments, as well as the remarks made by other teachers about the desire to see a variety of colors in the plot points, it became clear to us that a common attitude among the interviewees is that using color to interpret information in the prototype would allow for easier access to information and more accurate judgments about where individual students are located along the learning progression.

In addition to remarks about the visual representation of student scores, the interviews elicited some discussion about the display of information about the levels of the LP, specifically with regard to the About tab of the prototype and the level descriptions that are visible upon clicking the buttons in the whole-class and student views (see Figure 7).

Figure 7. Detailed information about LP levels in the About tab



Upon viewing the About tab, Deena expressed a positive opinion of the content of the page, but

also said, "I also wish there was like a comparison. So, this way you could see like all three levels on one screen, and all the different things they could do all on like one frame." Later in the interview, they expanded on this idea, stating that they would like to see

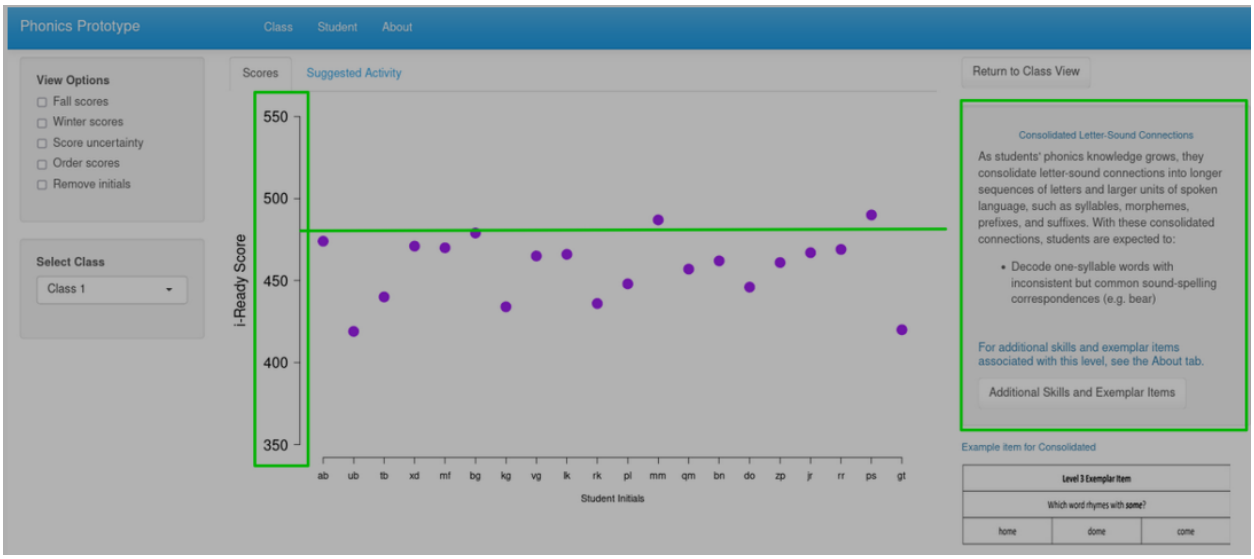
> maybe just a chart, or like, you know, it's got all three different levels, something like that. So then it's really all there in one document. So you could see like if somebody's doing partial and you could see where they need to be to complete. And then it'll give you a better picture of what you, as a teacher, need to do to get them up to the next level.

This seems to indicate that, for some teachers, clearly written content is necessary but not sufficient for an excellent user experience. Including options that allow teachers to view detailed information about the level of the LP in a compact view that allows for easy comparisons might enhance the overall experience. Amy shared a similar sentiment with regard to the Consolidated level view of the prototype.

> I'm not sure I totally understand this. I mean, I understand the concept of the Consolidated letter sounds, but I don't really know what the graph then is showing me. Even jumping back and forth, maybe if I saw them side by side, because it doesn't necessarily look like anything changed, and maybe it's not supposed to. The dots definitely look a little bit different … but I'm not exactly sure what I'm seeing on the graph.

This quote emphasizes the importance of presenting relevant information to teachers in such a way that does not require them to navigate multiple views. Amy seemed to be referring to the fact that when switching from the whole-class view to any one of the learning progression level views, the LP buttons and their corresponding reference lines on the plot disappear. Allowing the LP buttons, all of the horizontal LP level lines on the plot, as well as the level description to be viewed side-by-side might help teachers maintain their orientation when using this score reporting tool (see Figure 8 below for how these features currently appear in the prototype).

Figure 8. *Current layout of scale scores and horizontal line for a specific LP level description*



Theme 3: The phonics LP supporting PD

As per our third inductive theme, some teachers believed that the protocol could be a useful tool for professional development (PD). This is a topic that only emerged in a few interviews with teachers who were more experienced. For these educators, the prototype contained relevant information that is not always present in teacher training programs. As Amy explained,

> I think it's really good information. I think that it's very helpful, especially as there are so many teachers now that didn't grow up having any kind of phonics and not really having it in college either. So it's really helpful to have this so explicitly laid out, you know, as they are trying to learn this whole new way of teaching reading. And it's good reminders for me. I mean, I'm old enough that I had phonics in elementary school, but it's still good for me to know. You know, like specifically what one thing is called versus another, though I may have learned it, you know. They didn't teach explicitly, like we do now, using names of, you know, academic terms and that sort of thing. So this is, this is, I would utilize this a lot.

In this sense, the protocol could be a way to connect knowledge about phonics with real data about students in their classroom, which could in turn help teachers develop their own capacity in understanding and intervening at this level.

Table 3. *Count of Interviews in which each Inductive Theme was Identified*

	Positive		Negative	
Inductive Theme	Strong	Weak	Strong	Weak
Feedback on LP levels	0	0	1	4
Making sense of graphs	0	2	3	7
Use of prototype for professional development	1	0	0	0

Discussion and Future Directions

As was referenced above, we started this exercise with the experience that this research team had had with a similar prototype to showcase an LP for fractions based on a hypothetical fourth grade class (Briggs et al., 2023). Given that we interviewed in-service teachers in both cases, but about different content areas and topics (math/fractions versus reading/phonics), a valuable opportunity emerges to gather commonalities and reflect on how educators make sense of Content-Referenced Growth in different contexts. In our experience, it seems that educators who participated in the fractions prototype interviews were more successful in making inferences about student growth using the information that was presented to them in comparison to the educators that participated in the

phonics prototype interviews. For example, we found that in the fractions interviews, teachers more consistently made statements in which they noticed and interpreted how any given i-Ready scale score might correspond to a different level of student understanding of fractions. In contrast, as we described in the deductive and inductive themes sections of this report, teachers interacting with the phonics prototype tended to think aloud first about grade level proficiency expectations before considering student growth in terms of how children's knowledge of letters and sounds might be changing.

There are different possible explanations for this outcome. First, the content of reference of each prototype is very different. Fractions is a key and more discrete topic in mathematics that is nested as a specific concept within the numbers and operations strand of academic standards in math. Phonics, on the other hand, represents a much broader "big picture" concept around which the LP is oriented, with more skills, understandings, and common errors characterizing each level, as illustrated by the amount of text included in the About tab for the phonics prototype. Second, the educators who participated in the fractions prototype interviews taught upper elementary grades (3-5) and many of them were also instructional coaches, which suggests that they were both more specialized and may have been more experienced in teaching fractions, the content of reference. Teachers' reactions to the phonics prototype might be contextualized by a movement in elementary reading policy and practice driven by the Science of Reading (Shanahan, 2020), which has ushered in renewed attention to phonics and reading foundational skills, a focus that may not have been present in more seasoned educators' teacher preparation programs. Finally, the prototype that we used in the fractions versus phonics interviews was somewhat different: the fractions prototype was made in React and the phonics prototype was made in Shiny. While both platforms share many characteristics, React allows for more fine-tuned customization, while Shiny lends itself to relatively easy and rapid prototyping. There were some default settings in the Shiny app that generated differences in the user experience. For example,

as was referenced above, the Shiny app created a zoomed view when any level in the LP was clicked on which made it difficult to educators to visualize each level in the broad context of the LP.

These possible explanations point to key aspects of the experience of making sense of this prototype that we will focus on for future versions, motivating a number of potential updates and revisions we plan to pilot. One key aspect is that it is likely that different big picture concepts may require different levels of scaffolding for teachers. Whereas some LPs will be easier to understand for some educators (such as it seems was the case with the fractions LP), others might require more intentional effort in explanation before exposing educators to the data visualizations (such as it seems was the case with the phonics LP, especially given participating teachers' interest in more deeply understanding within-l). Related to this is the issue of the labels that are chosen to depict the levels of the LPs, which seem to be key for the educators' understanding of the LP levels. It seems that the labels were more effective in giving teachers the gist of the progression "at a glance" in the case of the fractions LP than in the phonics LP.

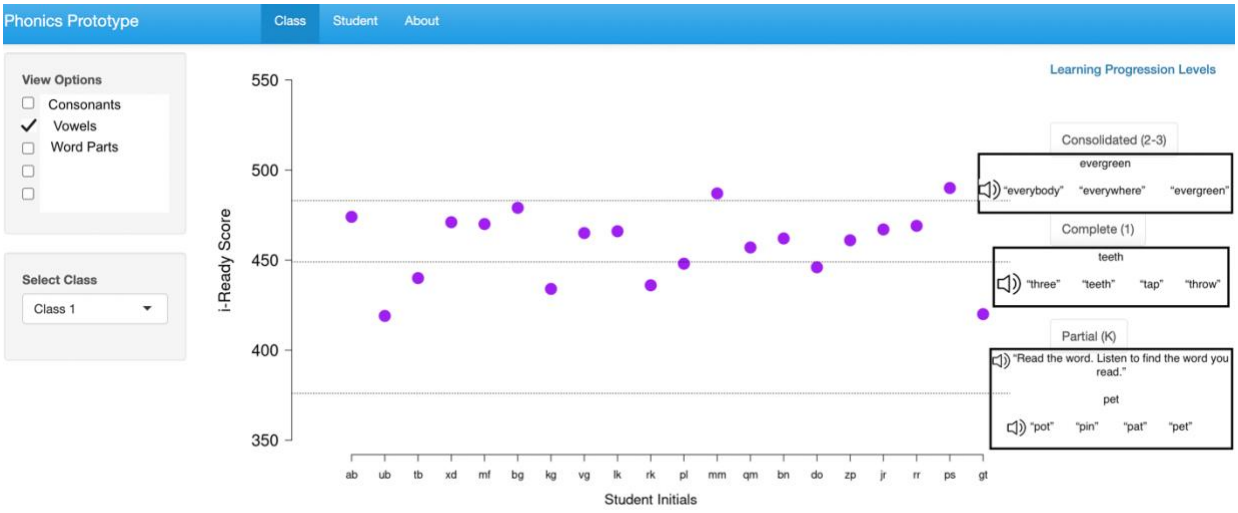
Another key aspect that we are keeping for future versions of this prototype is the importance of always showcasing the entire LP in the visualizations. Each of the levels is relevant in the context of the others, so visualizations that zoom in specific levels are less useful when they obscure the rest of the progression – in other words, when they focus more on the current status in the learning progression than in the growth within the progression.

Finally, an area we plan to explore is focusing more on specific exemplar items, rather than on descriptions of the different LP levels. One reason we see this as an exciting future direction is the feedback teachers shared in which they requested more information about the specific items students saw and answered on the Diagnostic assessment itself. Additionally, our hope for this prototype and the previous fractions prototype has always been that the qualitative distinctions between levels might be decisively communicated by high quality, highly illustrative exemplar items, without making level

descriptions overly text heavy. As such, using groups—or families—of exemplar items to emphasize the differences between levels is one revision to the prototype we see as important.

This would involve changing what appears as level descriptions in the Class tab. Specifically, we are considering creating checkboxes for three categories of letter-sound connections that extend across the three LP levels: consonants, vowels, and word parts. Once a user selects one of those categories, an exemplar item involving the respective category would appear under each level label to represent the kind of task about consonants, vowels, or word parts a student would be expected to complete at each level. As a very initial mock-up, we created the visual below in Figure 9.

Figure 9. Mock-up Visual of Item Families to Illustrate LP Levels



In this example, "Vowels" is selected as a view option, so there is an exemplar item displayed for each LP level where the target word for students to decode includes the vowel e. For the Partial level, the target word in the exemplar item is *pet*, reflecting the skill of decoding words with short vowels, which students are expected to be able to apply at this level. Next, at the Complete level, the exemplar item still involves the vowel e, but no longer the short e sound; rather, the target word, *teeth*, requires students to draw on their knowledge of the vowel team –ee as a way to spell the long e sound. Finally, the Consolidated exemplar item shows *evergreen* as the target word, a word that is multisyllabic,

contains an r-controlled vowel, but, just in the previous levels, includes the vowel e (with the vowel team –ee in the last syllable, *green*). One can imagine a similar chain in complexity of target words if the Consonants box were checked, perhaps from *slip* to *ships* to *shipwrecked* to show the transition students make from decoding words with consonant blends at the Partial level, to words with consonant digraphs at the Complete level, to multisyllabic words with consonant digraphs, prefixes, and suffixes at the Consolidated level.

This idea is preliminary and is one we will refine as we continue to reflect on the teacher interviews. However, it does represent an initial effort on our part to devise creative, effective ways for reducing our reliance on narrative text to explain the differences between LP levels to users of the prototype. There may be other strategies to achieve this goal, in addition to experimenting with the item family approach to distinguishing LP levels. We look forward to exploring additional possibilities as we continue to iterate on the prototype.

References

Briggs, D., Cox, O., Student, S., and Whitfield, E. (2023). Teacher Perspectives on the Content-Referenced Growth Reporting Prototype: Findings from Interviews. Center for Assessment, Design, Research and Evaluation.

Ehri, L. C. (2005). Development of Sight Word Reading: Phases and Findings. In M. J. Snowling & C. Hulme (Eds.), The Science of Reading: A Handbook (pp. 135–154). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757642.ch8

Clements, D. H. & Sarama, J. (2004) Learning trajectories in mathematics education, Mathematical Thinking and Learning, 6:2, 81-89.

Kennedy, B.L. and Thornberg, R. (2018) Deduction, Induction, and Abduction. In U. Flick (Ed.), The sage handbook of qualitative data collection. SAGE Publications Ltd, https://doi.org/10.4135/978152641607

The Reading League (2021). *Science of reading: Defining guide*. The Reading League. www.thereadingleague.org/what-is-the-science-of-reading

Snyder, C. M., & Delgado, H. P. (2019). Unlocking the Potential of Data-Driven Coaching: Child Assessment, Evidence as a Guide for Informing Instructional Practices. YC Young Children, 74(3), 44–53. https://www.jstor.org/stable/26789001

Appendix A. Teacher Interview Protocol

i-Ready Teacher Interview Protocol: Phonics Prototype

Fall 2023

Research Questions:

(1) How can a learning progression (LP) on phonics facilitate **meaningful connections** between scale scores on *the i-Ready Diagnostic* and the types of letter-sound correspondences students know?
(2) How are teachers **able to understand and interpret the learning progression** by using the prototype?
(3) How do teachers **envision using information** provided by the prototype?

Purpose: To solicit feedback from teachers about…

- the potential **usefulness** of embedding LP information into *i-Ready Diagnostic* score reporting.
- **substantive interpretations** supported by the prototype.
- the **prototype's usability**
  - Note: we expect Curriculum Associates to adjust/enhance the visual aspects of the prototype. However, our goal is to build out the functionality and information that is conveyed in the interactive tool.

Plan:

- Individual 1-hour long interviews with 8-10 teachers in grades K – 3 who teach reading.

Agenda:

- Introduction
- Teacher think-aloud for LP feedback
- Wrap Up

Intro Script:

Thank you for agreeing to join us! We're looking forward to getting your feedback about how we can make our work more useful for you, as a classroom teacher. We wanted to start with a brief overview of who we are, what we've been working on, and goals for what we are doing today. Please feel free to pause us at any point if you have questions or would like clarification. We'd also like to pause here to ask if it's okay for us to record this Zoom. We'll use the recording, along with our notes, to revisit ideas that come up in our conversation today. Is it okay for us to start recording?

(Also thank teachers who completed the DocuSign for informed consent OR ask if they need us to resend the form)

Who We Are

*First, we're going to introduce ourselves (*individually introduce who's on the call). *We are currently working in partnership with Curriculum Associates to help make their diagnostic assessment more useful and actionable for teachers.*

Goals for Today

Today we're going to show you a prototype of a reporting experience that our team is designing. The hope is that this experience makes assessment data more useful and actionable. We want to understand how users feel about the new experience, which is why we're getting feedback from educators like you. The reporting experience is structured around learning progressions. A learning progression describes how a student's understanding of a 'big picture' concept in reading becomes more sophisticated over time with the right curricular and instructional support. The big picture concept at the heart of our learning progression that we will be showing you today is knowledge of phonics, and the types of connections students form between letters and sounds. The prototype we've created connects this learning progression to students' performance on the i-Ready Diagnostic Assessment for Reading. We're going to ask you to interact with the prototype. This is an early draft so it is a great opportunity to provide us with feedback that can help shape the direction of the experience to be more useful and actionable for educators like you.

Question about current use of the *Diagnostic*

Given that goal, we are interested in knowing: How, if at all, do you currently use the i-Ready Diagnostic reporting system? How would you describe your--or your school's--instructional approach to teaching foundational reading skills like phonics? (Follow up: how frequently do you reference reports? How useful do you feel the reports are?)

Teacher Feedback Session:

Now we're going to have you interact with the prototype. Feel free to pause and let us know if you need to take a break for any reason or have questions at any point.

1) Start with LP-level view, no example item displayed, and no dropdown options selected.

Let's start with this. Could you start by just narrating what you're seeing and what you think it might mean or represent? (Give time for teacher to look around and narrate)

a. Prompts for think aloud narration: **What do you see? What does it make you think about? What do you want to do next?**

b. **How familiar are you with iReady scale score and how do you tend to interpret it? These buttons on the right, they're clickable they provide for another way to interpret the iReady scale.**

I'm going to open the prototype, then share my screen and give you mouse control through Zoom. (Use the mouse icon in the top bar of Zoom to give mouse control to participant). **Let's do a quick test to make sure you have control of the screen.**

Go ahead and click on anything that you see and want to explore. As you do, please narrate out loud what you are thinking. For example, you might say, "I'd like to be able to…" or "I wonder what this means" as you click on different features of the prototype.

(Provide 6-7 minutes for participant to click around on different tools/ functions)

Follow Up (If they ask about the horizontal lines and/or thresholds used to classify students by LP level)

"*The horizontal dotted lines you see connect the distinct levels of our phonics learning progression to specific thresholds on the i-Ready Diagnostic score scale. Fo example, the lowest line connects the partial letter-sound correspondence level with an i-Ready score of about 375. A student with this score is likely able to correctly answer most of the test questions that are associated with the partial letter-sound level of the learning progression*."

2) **How might you interpret the information presented here about this class of students?**

Follow Up Prompt (if teacher asks about score uncertainty range):

Every time a student takes this test, they take one set of questions. The score uncertainty range shows you how the student might have scored if they'd gotten a different set of questions. That's what we're trying to capture with this score uncertainty range. *The narrower the range, the greater the certainty.*

3) **How might you find more information about what a particular LP level indicates about students' phonics knowledge?**
    a. Name that you will be asking this question after giving them 10 minutes for them to read the gray level buttons and the About tab. Name that we'll be turning our cameras off, and they are welcome to do that too. **How would you interpret the differences between the levels of the phonics LP?**

       Follow Up Prompt: Feel free to take your time and review the levels from the gray buttons.

       Follow Up Prompt: Can you reflect on how useful (or not) the bulleted list of skills, exemplar items, and follow up activities are in helping you to understand what each LP level means?

4) (If teachers have not already clicked into the "Student" tab). **Now let's look at a sample student. How do you think you might find more information on an individual student's performance?**

       Follow Up Prompt: What if you tried clicking on the "Student" tab at the top?

a) **From the data you see here, how would you describe this student's phonics knowledge?**
b) **How would you describe this student's growth over time?**

5) **Where, if at all, does this kind of information fit within your instructional planning? What decisions would it help inform? How do you currently go about that? [In your instructional planning or practice, how might you use the information from this class-level view, if at all?]**

6) **If you were going to tell a colleague about what we looked at together today, what would you say?**

Wrap up:

Now that you've had a chance to see the entire prototype, we would love some of your general feedback. (Select 2-4 of the questions below, depending on time)

7) **Was there anything that you particularly liked or didn't like about the tool?**

8) **Is there anything you would like to see that is missing?**

9) **What questions do you still have about what we looked at together today?**

10) **What advice do you have for us as we continue to iterate on this work?**

11) **Is there any more information you feel like you need to understand the phonics learning progression better?**

12) **Knowing what we've told you about our plan for this project and what you've seen today, is there anything else you'd like to share with us? Is there anything you think we should have asked you about that we didn't?**

That wraps up our initial user test. You should be receiving your $75 honorarium in the form of an Amazon e-gift card from Sabine McKenzie, Curriculum Associates' User Experience Operations Manager. We'll put her email address in the chat, so you have it to reach out if necessary. We also wanted to ask: Would you be interested in hearing from us with updates about the progress of this project?

Thanks so much for participating today! We really value your feedback.