Assessing the Initial Implementation of New K–2 Assessments in Arkansas

Jeffrey Capizzano, Kelly Etter, PhD, and Soumya Bhat
The Policy Equity Group, LLC

THE POLICY EQUITY GROUP
EMPOWERING THE SOCIALLY CONSCIOUS

INTRODUCTION

In 2017, the Arkansas General Assembly passed Act 930, requiring public elementary schools to implement a developmentally appropriate literacy and mathematics assessment in kindergarten through second grade. To support implementation of the new requirement, the Arkansas Department of Education (ADE) issued a Request for Qualification (RFQ) in the spring of 2017 to assess various K–2 adaptive assessment systems. The RFQ process resulted in the selection of three approved K–2 assessments from which Arkansas school districts could choose in order to meet the assessment requirement. These new assessment systems were: (1) *Istation's Indicators of Progress (ISIPTM)*, (2) *NWEA's MAP® Growth™*, and (3) *Renaissance Star 360®*. Implementation of the new assessments began at the start of the 2017 school year.

Prior to the passage of Act 930, the state required the implementation of one screening tool at kindergarten entry—the *QUALLS Early Learning Inventory (QELI)*—that was used throughout the state. The QELI was a paper-and-pencil assessment completed by teachers that measured multiple domains of child development and learning, including general knowledge, oral communication, written language, math concepts, and attentive behavior. Kindergarten teachers and elementary school administrators cited numerous issues with the screener, including poor psychometric properties, burdensome implementation, and poor reporting functionality. These issues led the state to move from the point-in-time, paper-and-pencil-based kindergarten entry screener to the three computer-based K–2 adaptive assessment systems.

K–2 assessment tools

Arkansas's three new K–2 adaptive assessments represent a significant departure from the *QELI* in terms of the overall goals of the assessment, the domains that are measured, the way in which assessment data is collected, and reporting functionality. Unlike the *QELI*, the new assessments measure child growth over time and focus on only two aspects of development and learning—literacy and math. All three assessments are "adaptive," meaning that task/question difficulty changes throughout the assessment based on the child's success on previous tasks/questions. This assessment characteristic allows for more accurate assessment results, particularly for children who may be very low or very high performers. All three assessments include reporting functionality that allows teachers to understand a child's developmental level and growth shortly after the assessment has been conducted. Brief descriptions of each of the three approved assessments tools are provided in **Appendix A**.

Purpose of the report

This report, funded by the W.K. Kellogg Foundation, provides information about the initial implementation of the three new K–2 assessment systems. The report has three goals:

- To better understand school administrator and teacher perspectives regarding the transition from the QELI to the new K–2 assessments;
- To uncover potential implementation or measurement issues with the new K–2 assessments; and
- To gather information regarding additional resources that could support the implementation of the new assessments.

The findings and recommendations from the report can be used by ADE to determine whether additional guidance or implementation supports to the districts are necessary.

METHODOLOGY

To understand the implementation of the new K–2 assessments, survey data collected by ADE was analyzed and focus groups were conducted to assess the transition to and implementation of the three new K–2 assessment systems.

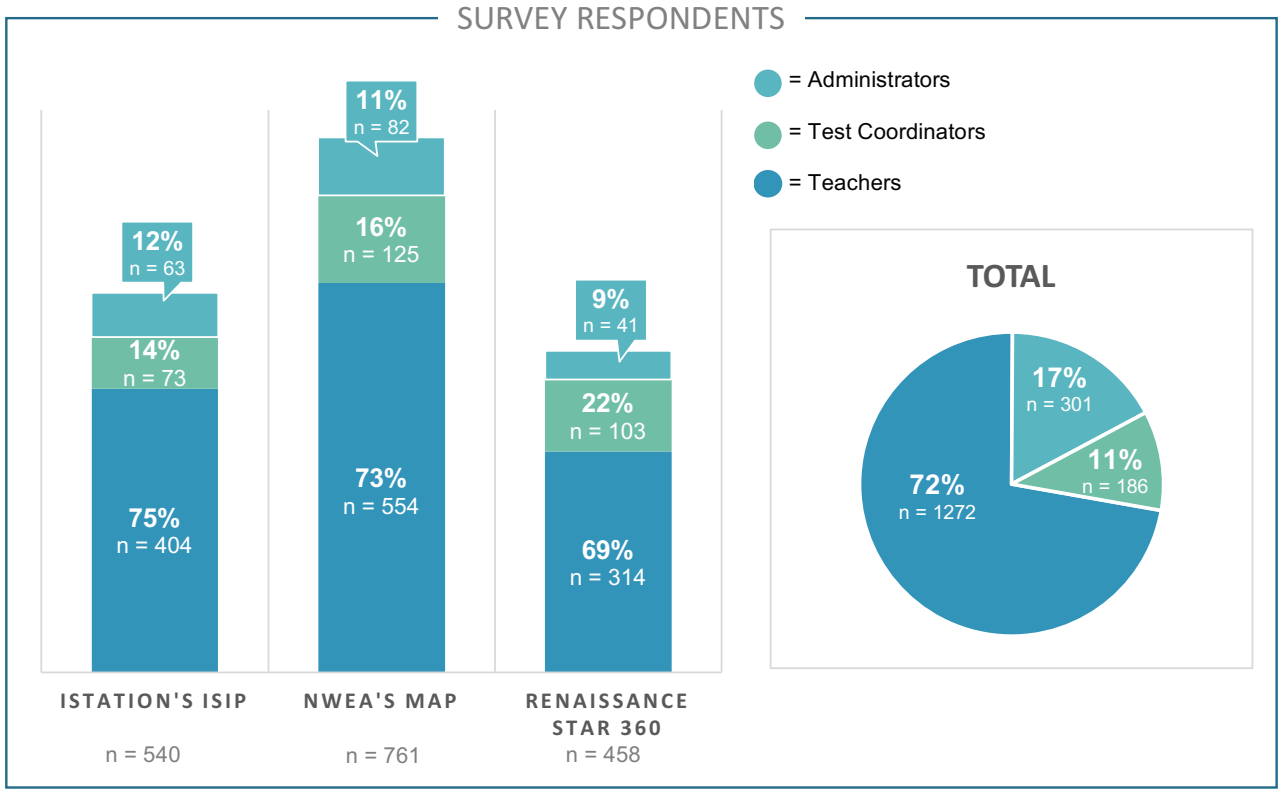
Survey

After districts began implementing the new assessment systems in the fall of 2017, ADE fielded a survey to gather information about administrator and teacher experiences with them. The survey included questions about the level of satisfaction with the training provided, the ease of use of the new assessment systems, the quality of customer service provided by assessment vendors, students'

level of engagement during the new assessment process, and other implementation-related questions. **Appendix B** contains the survey questions.

A total of 1,759 elementary school teachers, administrators, and other staff responded to the survey between October 10 and November 14, 2017. There was good representation from teachers, administrators, and test coordinators using each of the three assessment systems. **Figure 1** provides information on respondent roles by assessment. Thirty-one percent of respondents were using the Istation's ISIP, 43 percent were using *NWEA's MAP*, and 26 percent were using *Renaissance Star 360*. Across all assessments, teachers represented the largest share of the respondents, followed by test coordinators, and then administrators. Findings from the survey are presented in the next section.

Figure 1. Survey Respondents by Role and Assessment Used



Focus groups

In addition to an analysis of the survey data, nine focus groups were conducted between November 2017 and February 2018 to obtain more in-depth input on the transition from the *QELI* and the initial implementation of the new K–2 assessments. Six of the focus groups were conducted in person and three were conducted over the phone. A focus group protocol was developed to guide the focus group discussions, which included questions about the initial implementation of the tools, publisher supports, measurement issues, reporting functionality, and additional implementation supports that could be provided from ADE. **Appendix C** contains the focus group questions.

A total of 45 teachers and administrators across 11 schools participated in the nine focus groups. **Table 1** provides details on the focus groups.

Table 1. Focus Group Sites, Dates, Participants, and Data Collection Mode by Assessment

Assessment	District	Date	Participants	Mode
Istation's ISIP	Russellville	11/06/17	7	In-person
	Jasper	11/06/17	2	In-person
	Searcy	11/07/17	3	In-person
	Carlisle	11/07/17	4	In-person
NWEA's MAP	Hot Springs	11/14/17	6	In-person
	Springdale (3 schools)	1/25/18	7	Phone
Renaissance Star 360	Paragould	2/01/18	8	Phone
	Gurden	2/08/18	5	Phone
	Risen	11/07/17	3	In-person
Totals			45	9

Putting the findings in context

There are number of things to remember when reviewing the findings in this report. First, the survey consists of a self-selected sample of respondents. The survey was widely disseminated across the state to district test coordinators, and elementary school staff involved with assessment implementation decided whether to respond. As such, the pool of survey respondents may consist of individuals who feel strongly about the assessment process and may not be a representative sample of all teachers, administrators, and test coordinators across the state. Although the sample size for the survey is large and represents a broad number of teachers across the state, the findings should be interpreted as being from a sample of the state's more engaged teachers, test coordinators, and administrators.

Second, the focus group findings are based on a small sample of respondents. Although the findings may not be based on a representative sample, they represent the real experiences and opinions of the teachers, test coordinators, and administrators implementing the assessment. Common themes and issues were pulled from the focus group conversations and are presented in the findings section. The findings from the focus groups are remarkably similar to the survey results and provide a consistent story related to assessment implementation.

Finally, this report is not intended to assess the precision with which the assessment systems measure the reading and math abilities of children in kindergarten through second grade. Each assessment system has technical reports that provide information on the psychometric properties of the instruments. While questions were asked both in the survey and the focus groups that sought to uncover potential measurement issues, understanding how well these assessment tools measure the reading and math abilities of children in Arkansas will require a psychometric analysis of the data collected.

FINDINGS

The results of the survey and focus groups are discussed below.

Survey findings

The findings below are based on the data collected through the survey administered by ADE. Analyses of variance (ANOVAs) were used to examine differences in survey responses across the assessments systems and respondent type (i.e., teacher, test coordinator, or administrator). For questions that were asked of multiple groups of respondents (e.g., administrators and test coordinators), two-way ANOVAs were conducted, whereas for questions that targeted only one group (e.g., only teachers), one-way ANOVAs were used. Survey responses were examined for differences across assessments and respondent roles. Differences among groups were considered "statistically significant" if there was a low likelihood that the observed results occurred by chance (i.e., if the calculated probability, or p-value, was less than .05). If analyses indicated that there was a statistically significant difference across groups, follow-up post-hoc tests were conducted to examine specifically which groups were different from one another (i.e., which assessments differed, and/or which respondent roles were different). When relevant, analyses were also conducted to note if there were any significant interactions between respondent role and assessment (e.g., if teachers gave higher ratings than administrators on one assessment, but administrators gave higher ratings on another assessment). No significant interactions were found for any survey item.

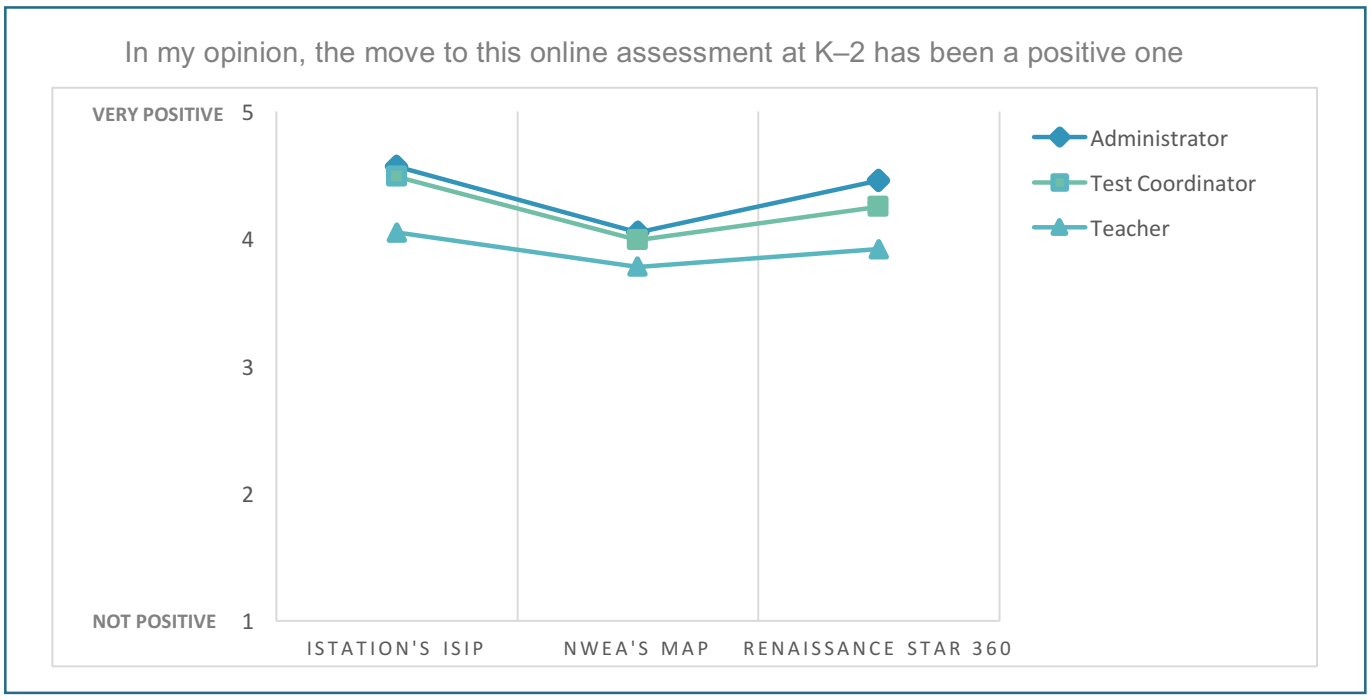
Perceptions of the Change to New Assessments

The first set of survey questions focused on respondent perceptions of the change from the *QELI* to a new assessment system. Overall, survey respondents reported that the change to the new assessment was positive, with an average rating of 4.02 on a scale from 1 (not positive) to 5 (very positive). This finding is not surprising given the significant issues reported with the *QELI* and the strong desire among teachers, test coordinators, and administrators to take a different approach to kindergarten entry assessment and assessment in the early grades.

As can be seen in **Figure 2** and **Table 2**, perceptions of the change varied by the assessment used ($F(2, 8) = 21.36$, $p < .001$)[1]. Specifically, respondents rated the change more positively for *Istation's ISIP and Renaissance Star 360* than for *NWEA's MAP* ($p < .001$ for both comparisons). There was no significant difference between *Istation's ISIP* and *Renaissance Star 360*.

Perceptions of the change also varied by respondent role ($F(2, 8) = 27.36$, $p < .001$). Administrators and test coordinators tended to rate the change more positively than teachers ($p < .001$ for both comparisons). There was no significant difference between administrators and test coordinators.

Figure 2. Perceptions of the Change to New Assessment



[1] For each *ANOVA*, the *F*-value is an indicator of how much individuals differ across groups compared to how much individuals in the same group differ from one another. Larger *F*-values indicate greater differences across groups. The *p*-value is the probability of getting that *F*-value or higher by chance rather than as a representation of true differences among groups (e.g., $p < .05$ means that there is only a 5 percent chance that the current data are an anomaly and do not represent a true difference in groups among the population of interest). If the observed *F*-value is larger than a standardized *F*-value based on the sample size and number of groups in the current analysis (represented by the degrees of freedom reported in parentheses after the *F*-value), it can be concluded that there is a "statistically significant difference" among groups.

Table 2. Perceptions of the Change to the New Assessment

	Istation's ISIP	*NWEA's MAP*	*Renaissance Star 360*	TOTAL
Administrator	4.57 (0.59) n = 63	4.05 (1.02) n = 82	4.46 (0.71) n = 41	**4.32** (0.86) n = 186
Test Coordinator	4.49 (0.58) n = 73	3.99 (0.94) n = 125	4.38 (0.79) n = 103	**4.25** (0.84) n = 301
Teacher	4.05 (0.91) n = 404	3.78 (1.00) n = 554	4.01 (0.92) n = 314	**3.92** (0.96) n = 1,272
TOTAL	**4.17** (0.87) n = 540	**3.84** (0.99) n = 761	**4.13** (0.89) n = 458	**4.02** (0.94) n = 1,759

Perceptions of Teacher Support for New Assessments

Administrators and test coordinators were asked about the supports provided to teachers and reported that teacher support of the new assessments has been strong, with an average rating of 4.07 on a scale from 1 (compliance only) to 5 (actively support). As can be seen in **Figure 3** and **Table 3**, there was a significant main effect for the assessment used ($F(2, 5) = 9.48$, $p < .001$). Specifically, administrators and test coordinators perceived teacher support to be stronger for *Istation's ISIP* and *Renaissance Star 360* compared to *NWEA's MAP* ($p < .001$ and $p = .02$, respectively). There was no significant difference between *Istation's ISIP* and *Renaissance Star 360*.

In addition, there was no significant difference between administrators and test coordinators in their perceptions of teacher support.

Figure 3. Perceptions of Teacher Support for New Assessments

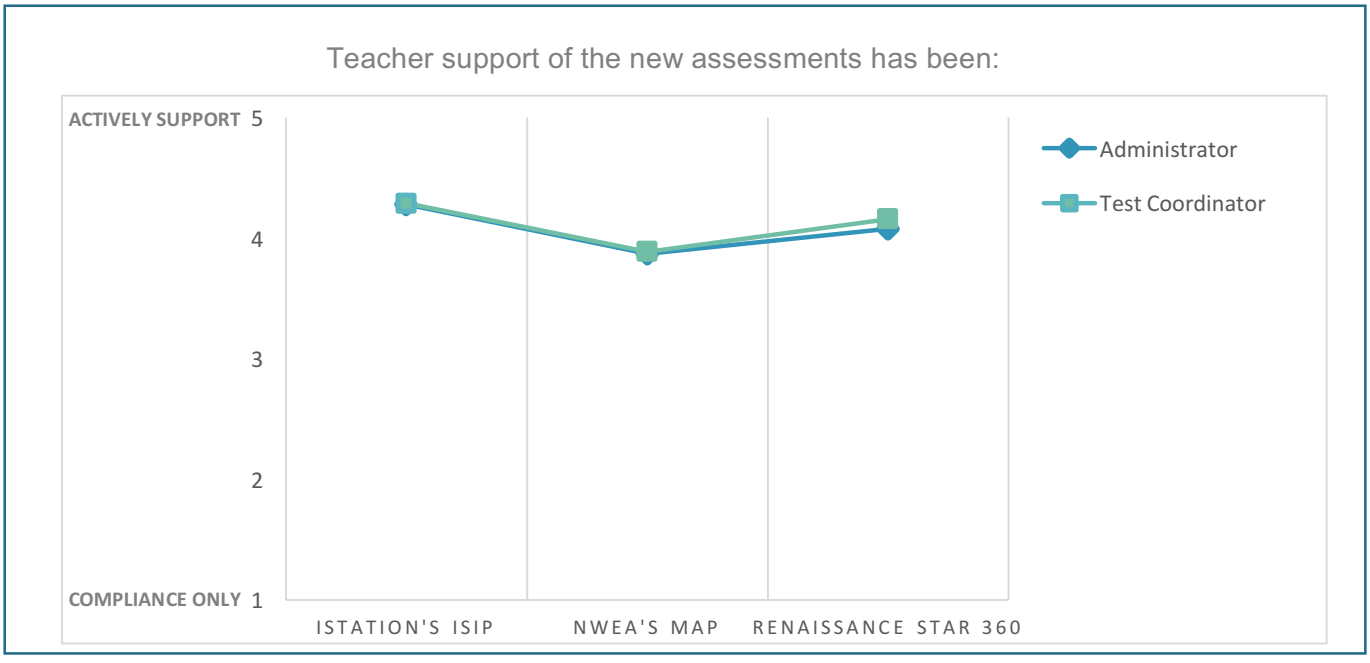


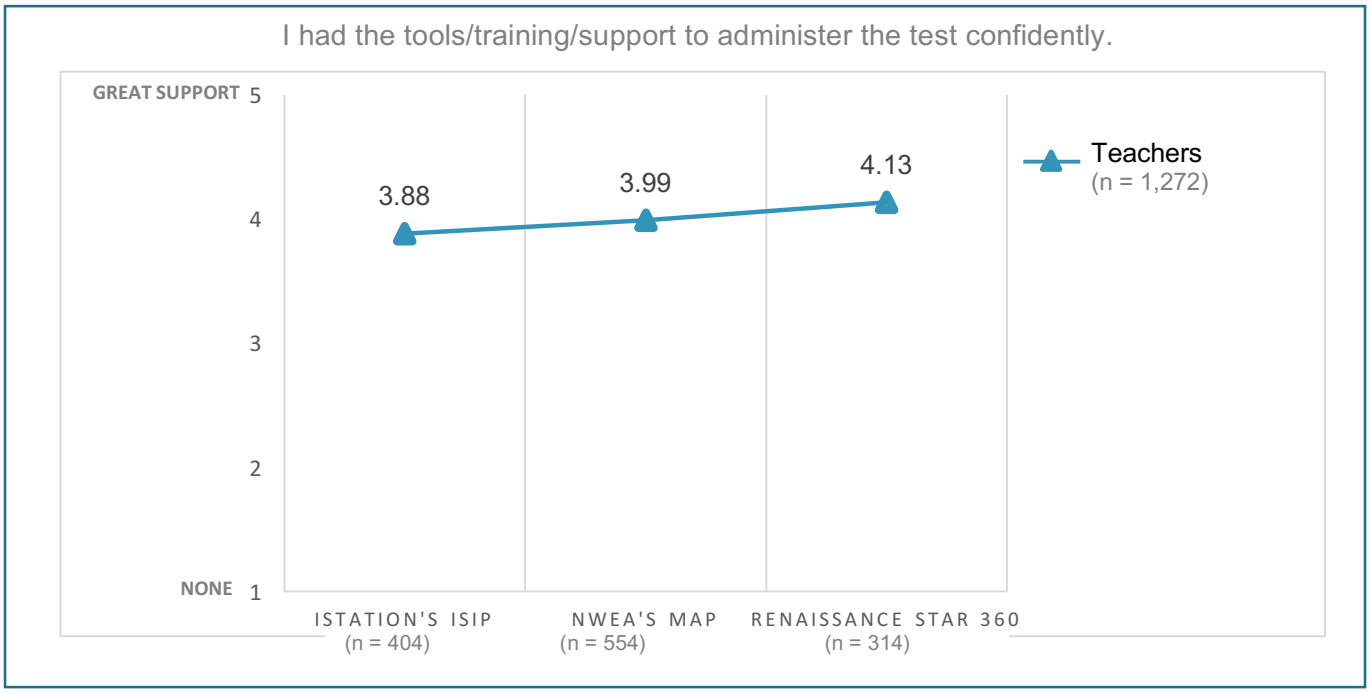
Table 3. Perceptions of Teacher Support for New Assessments

	Istation's ISIP	*NWEA's MAP*	*Renaissance Star 360*	TOTAL
Administrator	4.28 (0.69) n = 61	3.87 (0.97) n = 82	4.08 (0.89) n = 40	**4.05** (0.88) n = 183
Test Coordinator	4.29 (0.78) n = 72	3.89 (0.83) n = 124	4.16 (0.90) n = 102	**4.08** (0.86) n = 298
TOTAL	**4.17** (0.87) n = 133	**3.84** (0.99) n = 206	**4.13** (0.89) n = 142	**4.02** (0.94) n = 481

Teacher Ratings of Support Provided

Teachers were asked if they were provided the "tools/training/support to administer the test confidently." Overall, teachers reported that they had relatively high levels of support to administer the assessments, with an average rating of 3.99 on a scale from 1 (none) to 5 (great support). As can be seen in **Figure 4**, there were significant differences depending on which assessment teachers used ($F(1, 2) = 5.18$, $p = .006$). Specifically, teachers using *Renaissance Star 360* reported significantly higher levels of support than those using *Istation's ISIP* ($p = .004$). There were no significant differences between *NWEA's MAP* and *Istation's ISIP* or *Renaissance Star 360*.

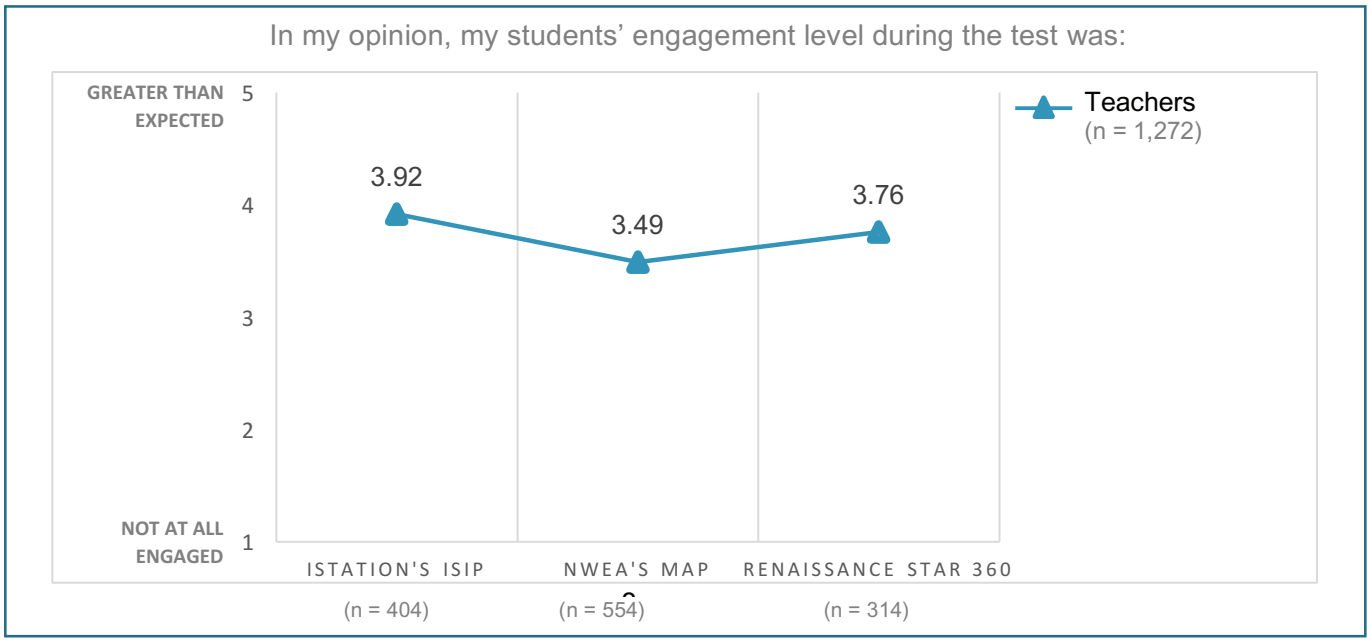
Figure 4. Teacher Ratings of Support Provided



Teacher Ratings of Student Engagement

Since the assessment systems are basically self-administering after some initial set-up by the teachers, it is important that students are engaged during the assessment in order to collect accurate data regarding their abilities. Accordingly, the survey asked specifically about student engagement with the assessment. Overall, teachers reported that their students were engaged in the assessment slightly more than expected, with an average rating of 3.69 on a scale from 1 (not at all engaged) to 5 (greater than expected). As can be seen in **Figure 5**, there were significant differences depending on which assessment teachers used ($F(1, 2) = 30.41$, $p < .001$). Specifically, teachers using *Istation's ISIP* reported greater student engagement than those using *NWEA's MAP* or *Renaissance Star 360* ($p < .001$ and $p = .036$, respectively).

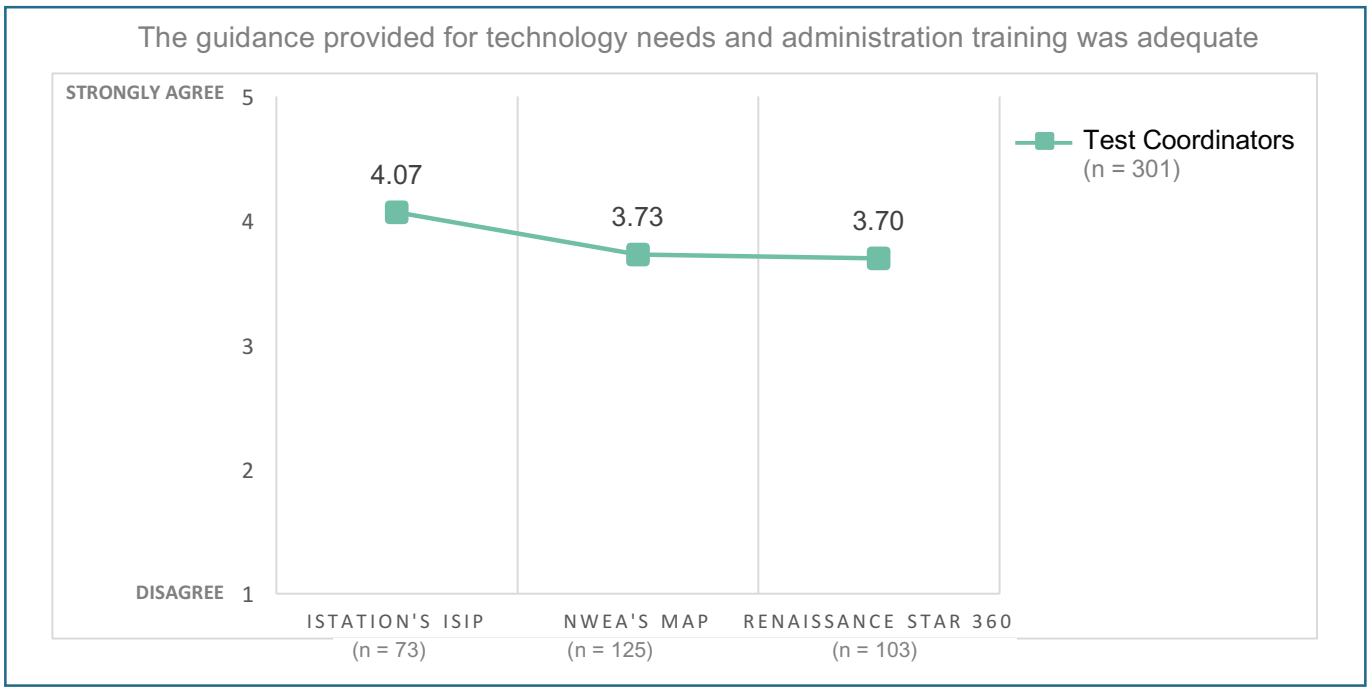
Figure 5. Teacher Ratings of Student Engagement



Test Coordinator Ratings of Guidance and Training

The survey also asked test coordinators whether the guidance provided for the technology was adequate. Overall, test coordinators reported a moderate level of agreement that guidance and training had been adequate, with an average rating of 3.80 on a scale from 1 (disagree) to 5 (strongly disagree). As can be seen in **Figure 6**, there were significant differences depending on which assessment was used ($F(1, 2) = 30.49$, $p = .032$). Specifically, test coordinators using *Istation's ISIP* reported stronger agreement than those using Renaissance Star 360 ($p = .043$). There were no significant differences between *Istation's ISIP* and *NWEA's MAP* or between *NWEA's MAP* and *Renaissance Star 360*.

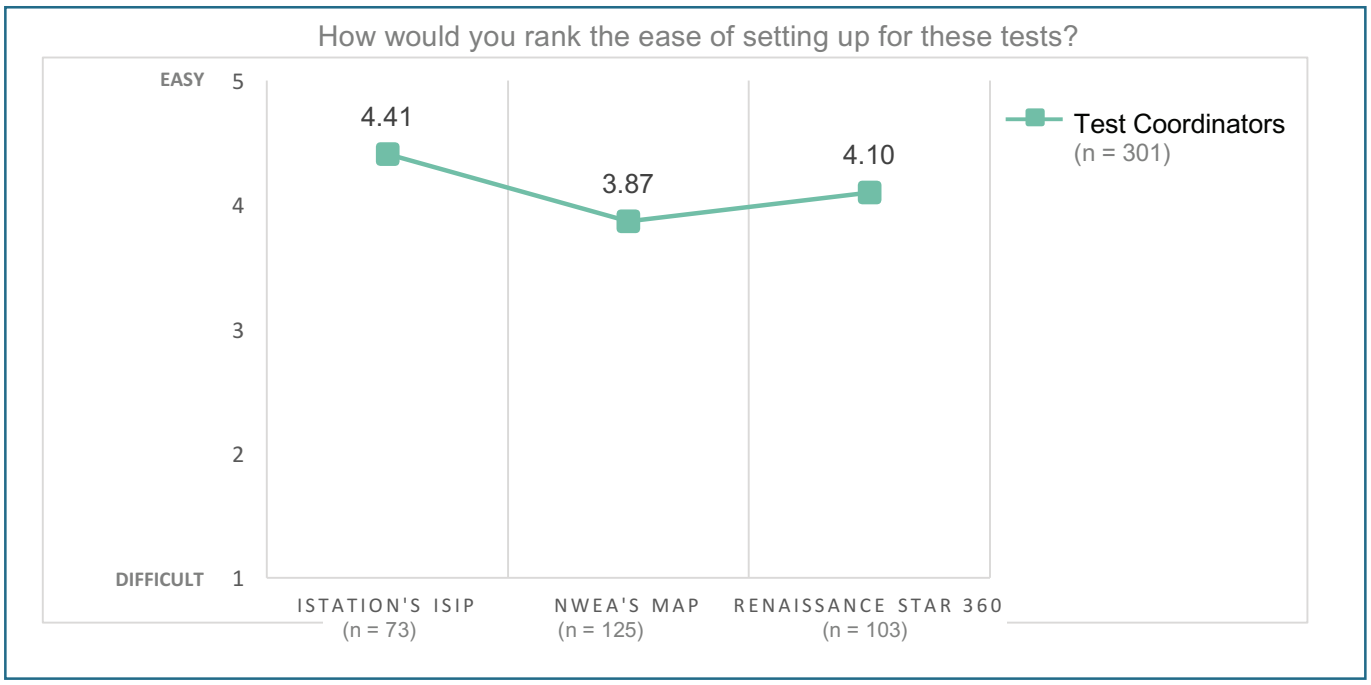
Figure 6. Test Coordinator Ratings of Guidance and Training



Test Coordinator Ratings of Ease of Set-up

The survey asked test coordinators about the set-up of the assessment systems. Overall, they reported that set-up was easy, with an average rating of 4.08 on a scale from 1 (difficult) to 5 (easy). As can be seen in **Figure 7**, there were significant differences depending on which assessment was used ($F(1, 2) = 8.46$, $p < .001$). Specifically, *Istation's ISIP* users reported easier set-up than those using *NWEA's MAP* ($p < .001$). *Renaissance Star 360* was not significantly different from either of the other systems.

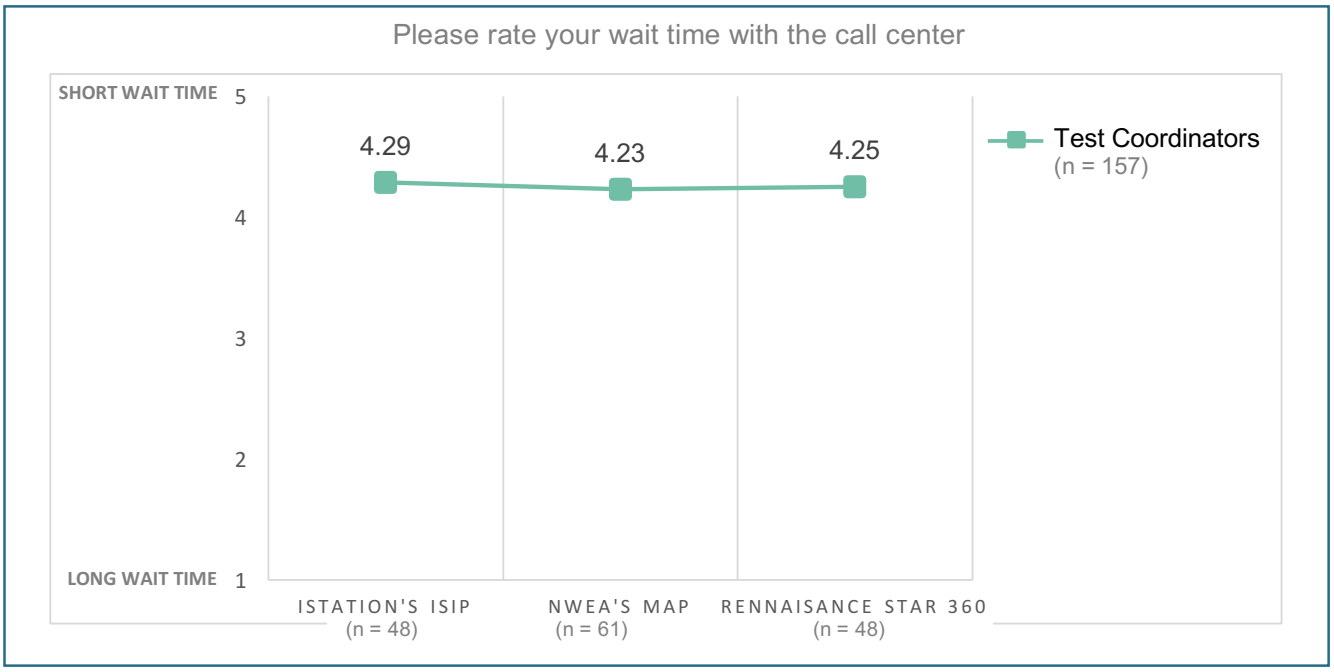
Figure 7. Test Coordinator Ratings of Ease of Set-up



Test Coordinator Ratings of Customer Service Wait Time

Overall, test coordinators reported relatively short wait times if they contacted customer service, with an average rating of 4.25 on a scale from 1 (long wait time) to 5 (short wait time). As can be seen in **Figure 8**, there were no significant differences among assessment systems.

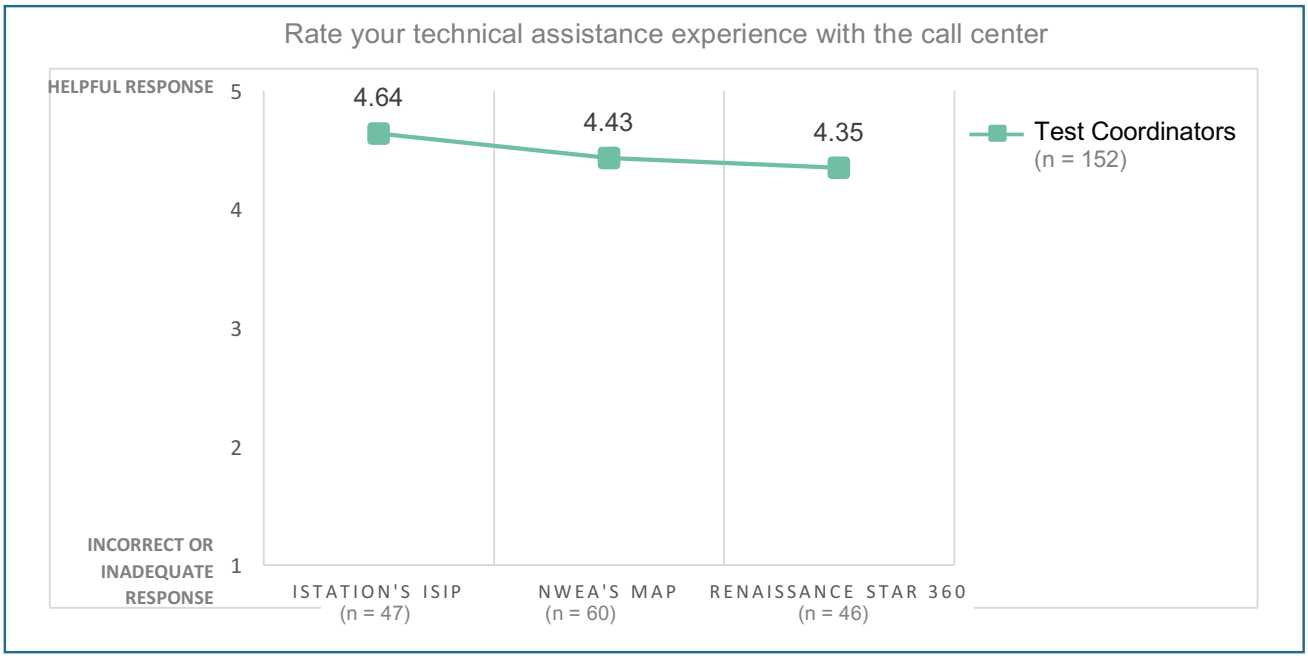
Figure 8. Test Coordinator Ratings of Customer Service Wait Time



Test Coordinator Ratings of Technical Assistance Experience with Call Center

Overall, test coordinators reported relatively helpful responses from the call centers if they contacted customer service, with an average rating of 4.47 on a scale from 1 (incorrect or inadequate response) to 5 (helpful response). As can be seen in **Figure 9**, there were no significant differences among assessments.

Figure 9. Test Coordinator Ratings of Technical Assistance Experiences



SUMMARY

Overall, the survey results indicate that the transition to the new assessment has been positive. Survey participants feel that they have been supported in the transition with adequate training and publisher support and feel that students have been adequately engaged in the new assessments. At the same time, teachers had the least positive reaction to the transition—lower than administrators and test coordinators—so it will be important to continue to support them in the transition and implementation. In addition, while ratings of ease of set-up were strong overall, test coordinators found some assessments less easy to set up than others. Assessment set-up was also a theme that emerged in the focus groups (discussed below), so it will be important to find easier ways of logging children into the assessment system.

Focus group findings

Focus group participants were asked a series of questions on topics similar to those addressed in the survey. The focus group format allowed participants to have a conversation around key transition and implementation issues and facilitated a richer understanding of the survey results. The focus group findings are consistent with those derived from survey data and are presented below.

Finding 1: Focus group participants found the K–2 assessment systems to be a significant improvement over the *QELI*

Consistent with the survey results indicating that teachers, test coordinators, and administrators found the change to the K–2 assessment systems to be positive, focus group participants noted that the new assessment systems were a significant improvement over the *QELI*. Participants cited two main reasons for the perceived improvement. First, the new assessment systems were significantly easier to administer. While the *QELI* required teachers to engage one-on-one with each child in order to administer the assessment, the new assessments were self-administering and allowed teachers/test coordinators to assess groups of children at one time. The only real limitation to the number of children assessed at one time was the number of iPads/laptops available to the teacher/test coordinator and the time it took to log children into the system. Notwithstanding these issues, the new assessments were significantly easier to administer than the *QELI*.

The second major reason cited for the improvement involved immediate access to the assessment data. Using the *QELI*, districts would send the assessment sheets to the testing company at the beginning of the school year for analysis and receive the results back weeks, if not months, later. This lag rendered the assessment data useless, as children progressed significantly over the time during which the results were being processed. Unlike the *QELI*, results from the new assessments are available immediately after administration. After children have been assessed, teachers can run a number of different reports at the class and child level, allowing them to individualize instruction, assign children to Response to Intervention (RTI) groups as necessary, and create other groupings according to the child's developmental level.

With the new assessment systems, teachers spent significantly less time assessing children yet obtained more meaningful and useful results.

Focus Group Participant Quotes on Administering the Assessment

"I've done the testing where you fill out the bubbles. Please God, don't go back to that. If we are going to have to do testing in kindergarten, this one is so much better. Having a teacher fill in bubbles is not appropriate."

"With Qualls you had to have it finished before you even knew the children; this gives you data to know the children." -Russellville teachers

"The stress level has been reduced for the teachers. It may be more accurate—the children are on the iPads and they can touch the answer rather than [the] teacher filling in bubbles." -Gurdon teacher

Finding 2: Overall, the transition to the new assessments went smoothly

Focus group questions investigated potential start-up challenges associated with the transition from the *QELI* to the new assessment systems. The new systems represent a significant departure from the *QELI* in the use of technology to conduct the assessment, the ways in which literacy and mathematics proficiency are measured, and reporting functionality. These changes require different teacher and administrator competencies and the development of new school-level assessment procedures. These changes, coupled with the fact that school districts had a small window of time in which to choose a new assessment, train staff on the assessment, and begin implementation, increased the probability of start-up challenges. Despite these circumstances, the transition to the new assessments went smoothly among the districts taking part in the focus groups.

The focus group discussions revealed that the districts took the decision of which assessment to choose very seriously. Key decision-makers within each district, which often included the superintendent, test coordinator, and a teacher representative, attended the vendor presentations. In some cases, scoring rubrics were created to judge the different assessment systems based on criteria most important to the district. In other cases, districts were already using the chosen assessment in a different capacity, which facilitated the decision-making process.

A key factor in facilitating the transition to the new assessment was the fact that the systems are self-administering. In other words, teachers worked to set children up on the new assessment system and then the program guided the child through the assessment questions. To take the assessment, a child wears headphones and listens to simple, age-appropriate instructions, and the intuitive nature of the assessment and technology allow the child to touch or click on a response to provide the answer. Teachers monitor children as they progress through the assessment questions. In comparison to the *QELI*, the technology takes on much of the burden of assessment implementation, making for an easier transition. This finding is consistent with the survey results.

Teachers in the focus groups also indicated that they received adequate training on how to use the new systems. In most cases, the districts participating in the focus groups used a train-the-trainer model where one administrator (e.g., the assistant principal) or test coordinator went through the publisher's training on the assessment and then trained district teachers. In some of the larger districts, webinars were developed and placed online in an effort to support teachers using the new assessments. In addition, specific district staff and the assessment publishers were available to provide support with problems that arose.

Although most of the feedback on the transition to the new assessment systems was positive, focus group participants noted some minor implementation issues. These issues included:

- iPads freezing up when logging on and when children were taking the assessment. This was due to a "compatibility issue" that was corrected by district technology staff and the publisher.
- Screen protectors on iPads made it more difficult for children to drag and drop as required by some of the assessments.
- Since the assessment systems rely on the internet, the speed of the school networks sometimes caused issues. In cases where the assessment was timed, slow networks would sometimes cause children to be logged out prematurely. Accordingly, it is important to have the appropriate internet speed for successful implementation.

Focus Group Participant Quotes on the Transition to the New Assessment Systems

"I'm pleased with the transition. It has assisted in our assessment process and teaching.... Kindergarten has never had testing where you can see progress—month by month, how kids are doing and we really like that.... When administrators talked about on-going assessment, they used to be like, 'kindergarten [teachers], you don't have to listen to this because it doesn't apply to you.' I need to see how far the kids have come. I need to see their progress. I need parents to see the progress. They make so many gains in kindergarten.... It's huge. With the Qualls, you filled it out and you were done."
-Jasper teacher

"This assessment gives specific information on how to help kids, where Qualls did not. Logistically, you spend one section of two days and then you are done. Whereas Qualls would drag on and on because of the manpower that it took and it would just go on and on. It is much less disruptive to a child's learning day." -Springdale teacher

"So far, it is pretty self-explanatory. When [children] are clicking…it basically lays in out for you." -Russellville teachers

By far, the biggest issue expressed by focus group participants was the time it took to log children into the assessment systems. Each child has a unique username and password, and in order to be sure that the assessment results are assigned to the right child, a teacher would need to enter the username and password of the child being assessed. This took a good deal of time and was considered an inconvenience more than an implementation issue. One district used an app that allowed children to log in by scanning a bar code specific to each child, which made the process simpler and more efficient.

Focus Group Participant Quotes on Logging into the New Assessment Systems

"It would take 15–30 minutes to get all of the students logged on and get their headphones on…. By the time you get your last student logged on, your first student is finished. You have to type in every user name and password." -Jasper teacher

Finding 3: Assessment publishers effectively supported implementation

When asked about publisher supports, focus group participants responded favorably. In most cases, districts had internal processes to support implementation and had established one point of contact to communicate with the publisher when necessary. Publishers' help desks were responsive to questions. These findings are consistent with the survey results related to the call center.

Finding 4: The adaptive nature of the assessment is a key strength

Focus group participants talked about the adaptive nature of the tests as a key strength of the new assessment systems. Adaptive assessments adjust the questions that are asked based on a child's proficiency and allow both high-achieving and low-proficiency students to receive accurate scores. Both teachers and administrators pointed to this characteristic as a major benefit of the new assessments. Focus group participants emphasized the fact that this information allowed them to "personalize" learning for all children in the classroom, even those at the highest and lowest levels.

Finding 5: There were some implementation issues that may impact measurement

Measurement Issue 1: Exposure to technology

Focus group participants indicated that the different technologies used to administer the assessment (e.g., iPad/Chromebook) caused some measurement issues. Some districts taking part in the focus groups were using iPads to administer the assessment while others were using touch-screen Chromebooks or other computers. Focus group participants felt that the intuitive nature of the iPad made implementation easier than trying to move a cursor on the screen to answer questions on the Chromebook. Moreover, in many cases children had little to no exposure to these technologies before entering school. Those children who lacked prior exposure to a computer or iPad had more difficulty staying focused, and were more interested in the functionality of the technology instead of the assessment questions. Teachers felt some initial assessments were not accurate as a result.

Focus Group Participant Quotes on the Exposure to Technology

"Some children have never seen a computer before.... Once children are logged on, they are curious and some will sit there and just punch buttons the whole time. You have to really keep them focused. If they have never seen a computer before and this is their assessment, we have to keep them focused the whole time." -Jasper teacher

"There is a little bit of pressure in kindergarten. The kindergarten students are the wonderers and thinkers so sometimes they get timed out and that is the number one thing I see on my reports." -Searcy teacher

Measurement Issue 2: English-only assessment

As noted in Arkansas's *Every Student Succeeds Act (ESSA)* plan, Ark. Code Ann. § 1-4-117 (2017) establishes English as the "official language of the state of Arkansas" and Ark. Code Ann. § 6-16-104 (2017) requires that "the basic language of instruction in … all the schools of the state, public and private, shall be the English language only." Accordingly, while translations of the assessment instruments exist, only the English version of the assessment can be used. Given this English-only assessment requirement, focus group participants were convinced that English-language learners were not being assessed accurately. Because districts are required to assess students in English, children who are English-language learners may not understand testing instructions or may not be able to answer the questions. They would therefore receive a score that does not measure their ability in the domain being assessed and does not provide useful information to inform instruction. This measurement issue also has implications for placing children in an RTI level that may not be warranted.

Focus Group Participant Quotes on Measurement Issues with Non-Native English Speakers

"I have five children in my classroom that are on Tier III and out of those five, four are Hispanic. I have to dig more because I know it is not a direct reflection of what they do know. We have to be really careful about that." -Searcy teacher

"We know that the academic results are not going to be very accurate with what they actually know because of that language barrier. But again, that is the nice thing that we give it three times because we can see as their language acquisition grows, their academic skills grow as well." -Springdale teacher

Measurement Issue 3: Question wording

For some items, the length of questions seems to cause response error. For example, children are asked to identify a word that is the same as another word, but by the time they go through the three words on the list, the child may have forgotten what the first word was. As such, the question is more a test of working memory than vocabulary. In some cases, the sounds children were asked to identify were not pronounced in the same way that they are typically pronounced in Arkansas (e.g., words like *men*, *pen*, *hen*, and *again*) and children struggled with these questions.

Focus Group Participant Quotes on Question Wording

"It was a multi-step [question]. There were three or four steps before they could answer. That was really hard for them because by the time they got all the way through…[the question] they would be lost. It was just too many steps for them."

"Some of the scores, you know that the child is capable, but sometimes even my top students will just sit there and listen and not answer the question. These are five and six year olds so you have some where they are not as focused on some of the lengthy instruction."

Finding 6: The assessment results and reports are very helpful in understating each child's development, how they should be grouped, and how to work with each child

A major criticism of the *QELI* involved the lack of reporting functionality and the lag time between implementation of the assessment and receipt of the results. Along with ease of implementation, the reporting functionality was noted as one of the most significant improvements of the new systems over the *QELI*. Each assessment system allows for the generation of a report at the child, classroom, school, and district level right after the assessment has been implemented. The child- and classroom-level reports are particularly important in helping teachers understand each child's literacy and math proficiency and for grouping children. Focus group participants mentioned that both teachers and administrators are still working to understand the constructs of the assessment to better understand and interpret the reports.

Focus Group Participant Quotes on Assessment Reporting Functionality

"From my standpoint, it is almost overwhelming. I can spend a lot of time looking through the reports of one class. One report tells me something, the next one tells me more, and the next one tells me more…"

Finding 7: There were some differences in implementation by assessment instrument

While the transition to the new assessment systems was positive across all of the instruments, and most of the implementation challenges were common to all three assessment systems, there were some findings specific to each assessment system. The findings are discussed below. It should also be noted that some teachers may have had prior experience with NWEA or Renaissance assessment systems, both of which have had a historic presence in Arkansas schools.

Istation's ISIP

Teachers and administrators in the focus groups who were implementing Istation's ISIP discussed how children enjoyed the "game-like feel" of the assessment. Children did not appear burdened or bored by the assessment, making it easier for teachers to implement. The game-like nature of the assessment most likely accounts for the high level of child engagement reported by teachers using Istation's ISIP on the survey. Indeed, Istation's ISIP received the highest child engagement score of the three assessments. At the same time, teachers were concerned that there is not a lot of variety in the games and that children may grow tired of the assessment tasks.

Teachers also liked the customized instruction suggestions that the system produced according to assessment scores. Districts used the results and curriculum it generated as part of an RTI process—lower tiers indicated that more intervention time using the Istation curriculum is needed. For example:

- Children in Tier III receive 240 minutes of *Istation* intervention every two weeks
- Children in Tier II receive 140 minutes of *Istation* intervention every two weeks
- Children in Tier I receive core classroom instruction

Finally, focus group participants using *Istation's ISIP* wondered why results for the reading assessment were disaggregated into component parts (vocabulary, phonemic awareness, etc.), but only one composite score was provided in math.

Focus Group Participant Quotes on Istation's ISIP

"Kids just feel like they are playing a game."-Searcy teacher

"I think the big pull for Istation was that it was going to test the students but you had something to support it.… You had curriculum help, lessons, and built-in support."

NWEA's MAP

Focus group participants using *NWEA's MAP* were particularly vocal about the high quality of the reports that were generated by the system. The reports supported individual instruction, the grouping of children, and understanding a child's progression of skills. Focus group participants found the "quadrant" report particularly helpful, which allows teachers to understand proficiency level and progress at the same time by grouping children into "high-achieving, high-growth" and "low-achieving, high-growth" categories.

Focus group participants noted that it is difficult to keep children's attention over the course of the assessment. The assessment time is the longest of the three systems (45 minutes) and the assessment was not described as having a game-like feel.

Focus Group Participant Quote on NWEA's MAP

"Just having the children at 5 years old at the beginning of the year taking a 45-minute test and keep their attention on the iPad that long—that is pretty tough."

Renaissance Star 360

Focus group participants were positive about the short implementation time of the Renaissance Star 360 assessment. This assessment appears to be the most efficient of the three assessments, requiring the least amount of assessment time to generate results.

Focus Group Participant Quote on Renaissance Star 360

"It's user-friendly—it doesn't take a lot of time." -Gurdon teacher

RECOMMENDATIONS

Consider implementing one statewide assessment of development and learning at kindergarten entry in addition to the district-level assessments.

Perhaps the only advantage of the *QELI* over the new assessment systems is the fact that it provided one consistent measure of child development and learning at kindergarten entry across the state. It is important to note that a key tradeoff of giving districts the choice to implement one of three assessment systems is that there is no longer a statewide measure of child development and learning at kindergarten entry. Statewide kindergarten entry assessment data is valuable because it can be compared across districts and allow for data-driven decision-making at the state level, helping to develop appropriate training and technical assistance for early care and education providers and teachers and supports for children and families. The state should also consider using a statewide assessment of school readiness that can be compared across school districts on key domains aligned with the *Arkansas Child Development and Early Learning Standards*. To minimize burden, this can be done using a sample of children rather than assessing every child in the state. As an example, Maryland has a statewide measure of school readiness based on sampling which it uses to improve early childhood program quality and practice.

Encourage cross-district collaboration and information sharing through a community of practice.

School districts within the state have already developed their own training materials, supports, and processes for successful implementation of the assessments that could be beneficial to other districts. For example, a train-the-trainer session conducted by assistant principals in certain districts or the online trainings conducted in Springdale may be helpful to other districts. As such, the state should consider developing a community of practice to allow school districts to connect with one another about their experiences with implementation as well as share training materials.

Provide more training on the interpretation and use of assessment reports to support instruction.

Focus group respondents talked about the wealth of information that is gathered from the assessments and highlighted how the system's reports helped inform instruction and more accurately identify the skills that students need to work on. At the same time, many participants also felt that additional training would help to maximize the reporting functionality of each assessment, particularly when talking with parents about their child's progress. Because the assessment itself is self-administering, less training is required for teachers to implement it, and a stronger focus could be placed on how teachers use and interpret the assessment data to support student learning.

Consider assessing other domains of development at kindergarten entry, particularly social-emotional development.

Another key difference between the new assessment systems and the *QELI* is a focus on only two domains of learning—literacy and math. In contrast, the *QELI* measured multiple domains in addition to literacy and math, including general knowledge, oral communications, and attentive behaviors. Neither assessment is well-aligned with all of Arkansas's expectations for children at kindergarten entry, as articulated in the *Arkansas Child Development and Early Learning Standards*. In particular, the lack of assessment on social–emotional development is particularly notable given the research showing its importance in student success.

While focus group participants noted that they could recognize behavioral issues in the classroom, there did not appear to be a focus on the social–emotional development of children. Research shows the dramatic impact that early relationships and social interactions have on a child's academic outcomes, mental health, and the success of future relationships.[2] As such, a focus on social–emotional development beyond behavioral issues involves working with children to develop trusting relationships with others, emotional expression and empathy, self-awareness, and a sense of identity. This critical area of development should be addressed and measured in kindergarten and the early elementary classrooms across the state.

Clarify that districts can assess English-language learners in their home language outside of the state assessment windows.

As discussed in the findings section, assessing non-native English-speaking children in English creates an issue with accurate assessment results. Arkansas law and the approved ESSA plan are clear about the requirements for schools to test in English during specific state testing windows (September, January, and April). However, there is a need for the state to clarify the rules related to English-only testing, which allow assessments to be conducted in other languages outside of the state testing windows. Both Renaissance Star 360 and Istation's ISIP offer assessments and other resources in Spanish and could be used by educators to gather additional information on the skills of their English-language learner students who speak primarily Spanish at home. This would allow for more accurate assessments of reading and math proficiency for the purposes of planning instruction and supporting English language acquisition.

[2]National Scientific Council on the Developing Child (2004). Children's emotional development is built into the architecture of their brains: Working paper No. 2. Retrieved from: www.developingchild.harvard.edu.

APPENDIX A: DESCRIPTIONS OF THE ASSESSMENT SYSTEMS

Istation's ISIP. Istation's Indicators of Progress (ISIP™) measures student growth through computer-adaptive diagnostic and screening programs. The formative assessments take about 30 minutes to administer and can be conducted on a monthly basis. It also delivers personalized data profiles in three ways: priority reports to help teachers better identify students for small-group instruction or other supplemental lessons; student summary handouts that outline each student's performance, usage details, Lexile® measure, assessment percentile rank, assessment grade equivalency, and priority alerts; and classroom summary reports to assist with grouping students and tracking skill performance. The Istation assessment can be used with an adaptive curriculum that personalizes instruction for students and includes resources to help teachers customize instruction, including lesson plans and automated tools. Thirty-one percent of districts in Arkansas chose the ISIP assessment. More information about the assessment can be found here: https://www.istation.com/

NWEA's MAP Growth. The NWEA MAP® Growth™ is a computer/tablet-based assessment that provides a standardized score for reading, language usage, and math. The score provides a snapshot of a child's achievement level relative to a national norm and can be compared over time to measure academic growth. Teachers track growth through the school year and over multiple years. The interim assessments are designed to be measured up to three times per grade and take approximately 45 minutes per subject area. Thirty-six percent of districts in Arkansas chose the NWEA assessment. More information about NWEA's MAP Growth can be found here: https://www.nwea.org/map-growth/

Renaissance Star 360. Renaissance Star 360® is an interim and formative assessment that monitors progress and student growth toward mastering state-specific learning standards for reading, math, and early literacy. The assessments take about 20 minutes to administer and deliver customized data about each student. The system includes a mastery dashboard showing which skills students have mastered and which need more instructional time, and tracks student data from multiple sources to provide an overall measure of student learning. The diagnostic report provides individual assessment data to help teachers identify appropriate interventions for their students. Thirty-three percent of districts in Arkansas chose the *Renaissance Star 360* assessment. More information about it can be found here: https://www.renaissance.com/products/assessment/star-360/

APPENDIX B: K-2 ASSESSMENT SURVEY-FALL 2017

This survey is designed to gather feedback from educators about the administration and use of the new K-2 assessments in Arkansas school districts. Any personnel involved in K-2 assessments are invited to complete this anonymous survey. Depending upon your role, different questions will be asked. We appreciate your time.

* Required

Untitled Section

1. **Which K-2 assessment did your school administer? ***
   Mark only one oval.

   ◯ Istation - ISIP
   ◯ NWEA - MAP
   ◯ Renaissance - Star

2. **What is your role? ***
   Mark only one oval.

   ◯ Teacher *Skip to question 3.*
   ◯ Administrator *Skip to question 14.*
   ◯ Test coordinator (Building or District) *Skip to question 24.*

Teacher Questions

3. **In my opinion, the move to this online assessment at K-2 has been a positive one. ***
   Mark only one oval.

	1	2	3	4	5	
not positive	◯	◯	◯	◯	◯	very positive

4. **I had the tools/training/support needed to administer the test confidently. ***
   Mark only one oval.

	1	2	3	4	5	
none	◯	◯	◯	◯	◯	great support

5. **Considering any preparation you did in the classroom to prepare your students for this online assessment, which of the following resources did you utilize? Check all that apply** *
Check all that apply.

- [ ] Used practice tests/items provided by vendor in classroom
- [ ] Ensured students had adequate access to computer prior to test day
- [ ] Integrated the use of technology into center time within my classroom
- [ ] Utilized computer lab time to help familiarize students with item types/function of the computer
- [ ] Explained to students what they would experience on test day during class time prior to testing day
- [ ] Other:

6. **Logging students into the test was:** *
Mark only one oval.

	1	2	3	4	5	
harder than expected	◯	◯	◯	◯	◯	easier than expected

7. **In my opinion, my students' engagement level during the test was:** *
Mark only one oval.

	1	2	3	4	5	
not at all engaged	◯	◯	◯	◯	◯	greater than expected

8. **Have you used the data from the fall administration?** *
Mark only one oval.

- ◯ Yes *Skip to question 9.*
- ◯ No *Skip to question 13.*

Yes, I have used the data from the fall administration.

9. **In which of the following ways have you utilized the data? Choose all that apply. ***
Check all that apply.

- [ ] to identify students for RTI/IRI
- [ ] to modify instruction
- [ ] to identify needed interventions
- [ ] reviewed in PLC/team planning
- [ ] Other: _____

10. **Which of the following resources have you utilized to access and interpret your K-2 data? Choose all that apply? ***
Check all that apply.

- [ ] Assessment LiveBinder
- [ ] Score Report Interpretation Guide
- [ ] Training at the coop
- [ ] Training within your district or school
- [ ] Collaboration with coworkers
- [ ] Direction from your principal
- [ ] Direction from your Instructional Facilitator
- [ ] Other: _____

Skip to question 11.

Teacher Questions

11. **How do you plan to communicate the scores home to parents? ***
Check all that apply.

- [ ] Parent/Teacher conference
- [ ] Open House
- [ ] Letter home with student
- [ ] Mail letter home
- [ ] Email to parents
- [ ] Use of existing online portal
- [ ] We do not plan to share these test results with parents
- [ ] We are unsure at this point
- [ ] Other: _____

12. **What observations or comments do you have about this new K-2 system (students testing online and 3 times in the school year)? ***

Skip to question 36.

No, I have not used the data from the fall administration

13. **What is currently prohibiting you from using the data? (mark any that apply) ***
Check all that apply.

- [ ] Have not attended training yet
- [ ] Do not know how to access my data
- [ ] Do not intend to utilize the data
- [ ] Principal does not support or require the use of the data
- [ ] Data does not provide me with useful information about my students
- [ ] I have not had time to look through the reports
- [ ] Other:

Skip to question 11.

Administrator Questions

14. **The change to this online assessment at K-2 been a positive one in my opinion. ***
Mark only one oval.

	1	2	3	4	5	
not positive	◯	◯	◯	◯	◯	very positive

15. **Teacher support of the new assessments have been ***
Mark only one oval.

	1	2	3	4	5	
compliance only	◯	◯	◯	◯	◯	actively support

16. **What feedback have you received from your school personnel about this change? ***

17. **Which of the following resources have you used to get answers to questions you had about the test and its administration? ***
Check all that apply.

- [ ] LiveBinder
- [ ] Vendor call center
- [ ] Phone call to ADE
- [ ] Email to ADE
- [ ] Coop
- [ ] Collaboration with other districts
- [ ] Testing coordinator, District and Instructional Staff (Instructional Facilitators)
- [ ] I have not had my questions answered
- [ ] Other:

18. **Have you seen the data from the fall administration? ***
Mark only one oval.

- ◯ Yes
- ◯ No

19. **What is your plan to ensure that the data are reviewed and appropriately used to inform teaching and intervention? ***

20. **Are you aware of the IRI requirements and how these test results can be used to inform intervention decisions?** *

Mark only one oval.

- ◯ Yes
- ◯ No

21. **Check all of the following resources you have utilized to access and interpret the data. Choose all that apply:** *

Check all that apply.

- ☐ Assessment LiveBinder
- ☐ Score report interpretation guide
- ☐ Training within your district
- ☐ Training at the coop
- ☐ Collaboration with coworkers
- ☐ Direction from your leadership
- ☐ Other: _____

22. **How do you plan to communicate the scores home to parents?** *

Check all that apply.

- ☐ Parent/Teacher conference
- ☐ Open House
- ☐ Letter home with student
- ☐ Email to parents
- ☐ Mail letter home
- ☐ Use of existing online portal
- ☐ We do not plan to share these test results with parents
- ☐ Other: _____

23. **What observations or comments do you have about this new K-2 assessment system (students testing online and 3 times in the school year)?** *

_____

_____

_____

_____

Skip to question 36.

Testing Coordinator Questions

24. **In my opinion, the move to this online assessment at K-2 has been a positive one. ***
Mark only one oval.

	1	2	3	4	5	
not positive	◯	◯	◯	◯	◯	very positive

25. **Teacher support of the new assessments have been ***
Mark only one oval.

	1	2	3	4	5	
compliance only	◯	◯	◯	◯	◯	actively support

26. **What feedback have you received from your school personnel about this change? ***

27. **Were resources you needed easy to find? ***
Mark only one oval.

◯ Yes
◯ No

28. **Which of the following resources have you used to get answers to questions you had about the test and its administration? Check all that apply. ***
Check all that apply.

- [ ] Assessment LiveBinder
- [ ] Vendor call center
- [ ] Phone call to ADE
- [ ] Email to ADE
- [ ] Coop
- [ ] I have worked with fellow testing coordinators
- [ ] Collaboration within my district
- [ ] I have not had my questions answered
- [ ] Other: _____

29. **Have you contacted the vendor call center/customer service? ***
Mark only one oval.

- ◯ Yes
- ◯ No

30. **If YES, please rate your WAIT TIME experience with the call center.**
Mark only one oval.

	1	2	3	4	5	
long wait time	◯	◯	◯	◯	◯	short wait time

31. **If YES, rate your technical ASSISTANCE experience with the call center.**
Mark only one oval.

	1	2	3	4	5	
incorrect or inadequate rsponse	◯	◯	◯	◯	◯	helpful response

32. **The guidance provided for technology needs and administration training was adequate. ***
Mark only one oval.

	1	2	3	4	5	
Disagree	◯	◯	◯	◯	◯	Strongly agree

33. **How would you rank the ease of setting up for these tests? ***
Mark only one oval.

	1	2	3	4	5	
Difficult	◯	◯	◯	◯	◯	Easy

34. **Did you face any of the following technology challenges in getting ready for or administering these tests? ***
Check all that apply.

- ☐ Difficulty rostering students
- ☐ Difficulty with wifi/internet access during testing
- ☐ Setting up devices
- ☐ Students being unfamiliar with technology
- ☐ Did not have the proper device to student ratio
- ☐ Students unable to login or getting kicked out repetitively
- ☐ Test was slow loading or advancing to the next question
- ☐ Items did not display correctly on the screen
- ☐ Other: _____

35. **What observations or comments do you have about this new K-2 assessment system (students testing online and 3 times in the school year)? ***

_____

_____

_____

_____

_____

Skip to question 36.

Final Section

36. **What additional training do you need at this time to efficiently utilize the data or administer the test?**

_____

37. **What additional support can the ADE provide to assist with K-2 assessments?**

38. **What is your job in the district?**

39. **Please provide any feedback you would like shared with the testing company. This could be anything from features about the use of the test you liked to areas you would like to see changed or enhanced, or just general feedback.**

40. **Please add any additional comments:**

APPENDIX C: FOCUS GROUP PROTOCOL

Understanding the Implementation of Arkansas's Kindergarten-2nd Grade Assessments at Kindergarten Entry

Kindergarten Teacher Protocol

Introduction

Thank you for taking the time to talk with me today. As you know, the Arkansas Department of Education replaced the Qualls Kindergarten Entry Assessment, or "KEA," with one of three different kindergarten through Grade-2 assessments (*Istation's ISIP, NWEA*, or *Renaissance Star 360*). The Head Start Collaboration Office received a grant from the W.K. Kellogg Foundation to gather information about the implementation of the new assessment and what supports or resources might help to enhance implementation. As part of the project, we are talking with kindergarten teachers in Arkansas about their experience implementing the new assessment, and asking about supports that would make the assessment process stronger and less burdensome. We are also interested in the aspects of school readiness that you think are the most important to measure. The new assessments cover only literacy and math, and we want to discuss other domains of development that you would want to measure and whether you are finding other ways to measure these domains.

I know you have recently finished implementing the new assessment for the first time, so we wanted to talk with you while the process was still somewhat fresh in your minds. I know how busy you are, and I can't thank you enough for talking with me today.

Informed Consent

This conversation should last a little over one hour. The information we gather from you today will be combined with information from other groups of teachers and published as part of a KEA report that will include recommendations on the criteria for a new KEA. The report may use quotes from you, but we will not use your name. We also may call out some specific examples of successes and barriers from your district, but again, no names will be used in the report.

Are there any questions before we begin?

Overview

1) Tell me about your school and your kindergarten classroom. How many teachers are in a classroom? How many children are in a classroom? Does each classroom have or share a teaching assistant?

2) Which of the three assessments did you choose and why did you choose that particular assessment? (*We know—just confirming.*)

3) Can you take me through the process of how you implement *Istation/NWEA/Renaissance*?

    a) *Probes:* When do you start? How long does it take? What mode of delivery do you use (computer, tablet, other)?

    b) Can you talk about how easy or difficult it has been for the children to use the technology to implement the assessment?

    i) Do technological issues get in the way of accurately assessing the children? If there have been issues with technology, can you describe them? Do most children have some familiarity with the technology (computer/tablet) or is it new to many?

    ii) How are the teachers adjusting to using the computer/tablet to assess the children?

    iii) What is the hardest part of implementing the assessment? What is the easiest part?

    iv) How do you prepare children to take the assessment?

    v) What is the most engaging part of conducting the assessment? What is the least engaging?

4) How many other assessments are you doing at the same time? Do they work together? Do you feel that you can eliminate other assessments with the information you are getting from the new assessment? If so, which ones? If not, why not?

5) What do you like most about the new assessment?

6) What do you like least about the new assessment?

7) In what ways does the new assessment help you improve your teaching and interactions with individual or groups of children?

   a) *Probe:* Is this helping you plan instruction? Do you think these assessments will make a difference in children getting on a path to read at grade level by the end of third grade?

8) In what ways, if any, does the new assessment detract from your teaching and interactions with children or displace other valuable classroom activities?

Organizational and Professional Capacity

9) How were you trained to implement the new assessments? What additional tools or supports did you utilize to support implementation? Does your school have any tools or resources or trainers to support implementation?

   a) *Probes:* Is there a person you can go to for help? What supports would you need to better facilitate the assessment?

10) How helpful is the publisher of the tool in answering your questions or responding to your requests for support? Have you called their support line? If so, for what reason and what was the response or outcome?

11) Did the publisher of the tool provide enough support when you were initially making the transition to the tool? Can you describe the "onboarding process?" What went well and what didn't?

   a) How difficult has the transition been to the new assessment? What were the major stumbling blocks?

   b) What was the initial training like? Did it cover everything that you needed covered to successfully implement the assessment? What would you have added to the training?

12) How useful are the reports generated by the new assessment? Do you have support in analyzing and responding to the reports that are generated? How was the training related to reading the reports and using the data?

Implementing the KEA

13) Are there any issues with obtaining an assessment score for some children?

   a) Dual language learners

   b) Children with disabilities/cognitive delays

   c) Chronically absent children

   d) Others?

   i) Do you use any type of modifications (translation, additional time, or other modifications) to support children when implementing the assessment?

   ii) Do you feel the scores for these children are accurate? Why or why not?

14) What other issues (if any) come up that make it more difficult to successfully implement the new assessment?

15) Do you talk about the KEA process with parents? Do you share the results with them? How do you use/share the data with parents?

16) School readiness typically covers multiple domains of development. Are you doing any other assessments to measure other domains of development (like social¬–emotional development)?

   a) Do you anticipate that the additional assessment is giving you the data you need, or is there a better solution for collecting KEA data? What does this assessment look like?

   b) In an ideal world, what data would you want, how should it be presented, and when would you need it?

17) What advice would you give the state as they work to support implementation of your new assessment instrument? How can they be more supportive to you?

18) Are there other things the state should know about the new assessment that I did not ask you about?