CEMETR-2020-01
MAY– 2020

CEME Technical Report

The Center for Educational Measurement and Evaluation

Recommendations for the Implementation of a Statewide Formative Assessment Process

Angela M. Ferrara

Richard G. Lambert

T. Scott Holcomb

A PUBLICATION OF THE CENTER FOR EDUCATIONAL MEASUREMENT AND EVALUATION

CEME
The Center for Educational Measurement and Evaluation
College of Education
UNC CHARLOTTE

Recommendations for the Implementation of a Statewide Formative Assessment Process

Angela M. Ferrara, Richard G. Lambert, and T. Scott Holcomb

Center for Educational Measurement and Evaluation

The University of North Carolina at Charlotte

May 2020

Table of Contents

Recommendations for the Implementation of a Statewide Formative Assessment Process

Executive Summary

Early childhood assessments should be used to guide teaching and learning, identify students that may potentially need interventions, and improve educational programs (NAEYC & NAECS/SDE, 2003). Kindergarten Entry Assessments (KEA) are implemented at the start of kindergarten to provide a glimpse of where students stand at the beginning of the year across different developmental domains. The North Carolina Kindergarten Entry Assessment (NC KEA) can be a very helpful resource when teachers use it to get to know children at the beginning of the year. It can help teachers understand strengths that each child brings to the classroom and specific areas where each child needs support. The data that the NC KEA provides is intended to emerge from the ongoing processes through which teachers gather rich portfolios of evidence concerning student growth, analyze those evidences, make periodic placements on developmental progressions based on those evidences, and use those placements to plan and support the next steps in the learning process (Lambert, 2018).

This report contains information on the background of and research on the NC KEA. Evidence shows that implementing formative assessment is difficult for teachers and schools. There is distinct variability in teacher understanding of the purpose of the NC KEA, the process of implementation fidelity, and application of the information to classroom practices. Given that North Carolina will be adopting a new KEA measure for the fall of 2020, this report summarizes previous research on the NC KEA, as well as formative assessments throughout the United States, in order to provide a set of data-based recommendations to help support implementation during this transition. The recommendations provided in this report can be divided into three broad areas: initial and ongoing professional development, supporting implementation practices, and administrative best-practices. The following is a condensed list of these recommendations:

Initial and Ongoing Professional Development (PD)

For classroom level practitioners:

- Initial assessment specific training must extend beyond an introduction of its content and demonstrating the mechanics of the assessment software. It must also:
    - include instruction on the early childhood principals grounding the assessment,
    - provide a thorough explanation of each step in the formative assessment process cycle,
    - communicate clearly the purpose and its intended use at ALL implementation levels from the state down to the classroom, and
    - provide clarity around how the new assessment fits into the broader educational landscape (i.e. how it aligns with and supports other state and/or district standards, curricula, and assessment mandates).
- Ongoing professional development is necessary to overcome implementation barriers and ensure practitioners implement with fidelity. This PD should:
    - further develop teachers' understanding of how to make valid placements on developmental progressions,
    - provide in-depth training on the features of the electronic assessment platform that will assist teachers in determining next steps for their students from a whole-child perspective, and
    - train teachers to use the formative data the assessment provides to individualize instruction to support their students' growth and development.

For district and school administrators:

- Provide PD around early childhood development principals and the whole-child philosophy, best practices in early childhood instructional methods, and the foundations of formative assessment practices.
- Initial and ongoing PD should explicitly explain how the formative assessment aligns with other state and/or district mandates, most especially how the assessment can support student outcomes on accountability measures and provide updates in this alignment as mandates shift.

Supporting Implementation Practices

- Professional learning communities (PLCs) and instructional coaches/mentors are powerful resources that should be leveraged to support implementation.
- PLCs should allow time to discuss and address implementation barriers, troubleshoot software issues, collaboratively analyze collected evidences in order to improve teacher understanding of the progressions and reliability of their status placements, and make collective instructional decisions using assessment data.
- A coaching model leveraging existing instructional coaches/curriculum specialists can provide ongoing professional development through one-on-one mentoring sessions, as well as group support through grade level planning or PLC meetings.
- Mentors/coaches should utilize an implementation fidelity measure that assist them in identifying which steps of the formative assessment process practitioners may need additional support.
- Provide implementation support resources, which:

- help teachers determine next steps for students based on their placements on the learning progressions. This includes planning for individual, small group, and full class instruction.
- help teachers view the assessment data holistically to gain an understanding of the whole-child (i.e. how one domain may be affecting another in regard to their placements on those progressions).
- continually reinforce practitioner understanding of the purpose and value of the formative assessment, and how it fits into the 'big picture' of early childhood instruction in their state, district, and school.

Administrative Best-Practices

- Disseminate notifications quickly as assessment information and resources become available or are updated.
- Ask teachers how they can be supported with implementation and provide time, PD, and/or resources as necessary to meet those needs.
- Encourage discussions of formative assessment data in planning meetings to emphasize the assessment as a valuable tool and an instructional priority.
- Provide opportunities for implementers to meet with their peers to discuss the assessment process and how to apply the data to instruction.
- Further develop school-level implementation teams to assess current/future professional development needs, identify implementation barriers, and strategize solutions.
- Encourage district implementation teams to communicate openly with their regional implementation leads about their implementation progress and needs.

Conclusion

Teachers who received well-structured and thorough KEA specific professional development perceived a strong sense of efficacy in their utilization of the assessment as they felt they understood the content, its process, and its overall purpose. Therefore, with respect to future teacher training for the implementation of formative assessment processes, we recommend starting with clearly communicating the intended purpose of the assessment, including how the data is meant to be used at every implementation level, from the state down to the classroom. Next, teachers need to understand the process of collecting valid and representative evidences of student learning that align to the progressions. Next, teachers need to learn how to analyze evidences, place students on the progressions based on that evidence, and how to apply this data to their instruction. Throughout implementation teachers, schools, and districts need to have ongoing support for each of these components of the formative assessment process through ongoing professional development, direct coaching, professional learning communities, and online resources.

Administrators play an important role in supporting implementation in their districts and schools. A major theme across kindergarten teachers that have implemented the NC KEA was that it is more developmentally appropriate for students, but often framed as a lower priority than other assessments by their administrators. With any new formative assessment initiative, administrators need to understand its purpose, how it aligns with student learning standards, and how it fits into the broader educational landscape in the early grades. Administrators also need to be trained to help teachers understand how to use data from the assessment to guide student learning, including how to incorporate the assessment into data discussions during grade level planning, PLC, and vertical planning meetings. This is important to support implementation, as

studies have shown that where strong collaborative teams and PLCs exist, teachers exhibited stronger agency in the use of the NC KEA as intended and utilized data more frequently to inform their instruction.

For more information about each of these issues and more detailed recommendations, the full report is formatted in the following order: I. Introduction - Background of NC KEA, II. Research on the NC KEA, III. Previous research on formative assessment and KEA's across the United States, and IV. Preparing teachers to make valid placements on developmental progressions.

I. Introduction - Background of the North Carolina Kindergarten Entry Assessment

Kindergarten Entry Assessments (KEAs) are tools used at the beginning of kindergarten which can be used to gather data about the emergent academic, social-emotional, self-regulatory, cognitive, and physical development, knowledge, and skills of ingoing students. These data create a snapshot of the whole child developmentally, thus assisting educators in identifying and supporting students who enter formal schooling not meeting key readiness standards (Cohen-Vogel, 2011). The Obama Administration's Race to the Top-Early Learning Challenge (RTT-ELC) promoted the adoption of KEAs by providing competitive federal education grants to states in order to develop their own KEA programs. Since the RTT-ELC was launched in 2011, 30 states have enacted state-wide KEA policies. (Merrill, Cohen-Vogel, Sadler, Little, & Lee, 2017).

As an RTT-ELC grantee and in response to a North Carolina (NC) State Legislative mandate requiring the development of a kindergarten entry assessment, the Office of Early Learning (OEL) at the NC Department of Public Instruction (NCDPI) organized a K-3 Assessment Think Tank in February of 2013. That group, which included K-3 teachers, administrators, parents, early childhood development scholars, and policymakers, were tasked with researching a plan for developing effective, developmentally appropriate, student-centered assessments to improve early childhood learning and instruction in NC public schools. Their final report (NC K-3 Assessment Think Tank, 2013) was heavily influenced by prior research on the effectiveness of formative assessment on student learning outcomes (Black & William, 1998; William & Thompson, 2007), and recommended the development of a K-3 formative assessment designed to help teachers individualize their instruction. Their report further identified five domains of learning the assessment should address: (1) approaches to learning, (2) cognitive

development, (3) social and emotional development, (4) health and physical development, and (5) language development and communication.

Based on the recommendations of the K-3 Assessment Think Tank, an assessment design team consisting of current and former early childhood educators and scholars developed the North Carolina Kindergarten Entry Assessment (NC KEA). The NC KEA is intended to generate data about kindergarten students' abilities across the five focus areas of early childhood development outlined in the Think Tank report by providing teachers with a framework to observe their students during the course of their normal daily instruction in a more intentional way. Kindergarten teachers assess children on multiple developmental progressions, each identifying a fundamental skill which children generally acquire around the age of kindergarten entry. The assessment design team developed a total of 17 different developmental progressions. Of these, 10 were utilized as a part of the 2014 NC KEA pilot, three were implemented during the initial NC KEA statewide roll-out in the Fall of 2015, and eight have been implemented consistently over the last four academic years (2016-2019). These eight developmental progressions include: engagement in self-selected activities (Approaches to Learning), object counting (Cognitive Development), emotional literacy (Emotional-Social Development), letter naming and following directions, (Language Development & Communication), and grip and manipulation, crossing midline, hand dominance (Health & Physical Development).

Since the assessment's inception, researchers from the University of North Carolina at Charlotte Center for Educational Measurement and Evaluation (CEME) have conducted three separate qualitative research studies on the NC KEA. These studies were a strategic portion of North Carolina's overall implementation plan. Built on the implementation frameworks of Dean Fixsen and his research for the National Implementation Science Network (2013), OEL included

periodic improvement cycles informed by practitioner feedback. The purpose of CEME's studies, therefore, was to inform potential changes to the assessment content, online assessment platform, professional development, and implementation resources and support structures, and to better understand overall practitioner perceptions of the formative assessment process.

For the fall of 2020, North Carolina will be transitioning from the North Carolina Kindergarten Entry Assessment to the North Carolina Early Learning Inventory. This new measure is based on the Teaching Strategies GOLD Assessment System (GOLD). The connection to GOLD brings several significant advantages to our state. First, GOLD has a well-developed inter-rater reliability certification system. Second, GOLD offers a comprehensive set of progressions that cover all domains of child development. Third, GOLD provides "color bands", for each of 60 developmental progression, which define for teachers the range of skills and abilities that constitute "widely held expectations" for child development across age / grades birth to third grade. Fourth, GOLD provides teachers with specific suggestions for instruction, including small group and whole group activities, and targeted strategies for differentiation and individualization. All of these advantages can be tangible resources for teachers if teachers are trained and supported to implement the system to full fidelity. On the other hand, GOLD can be a significant demand on teachers if they are not equipped and supported to use it as intended. This report focuses on strategies that can help teachers fully realize all of the strengths of the GOLD assessment system.

II. Research on the North Carolina Kindergarten Entry Assessment

Research Phase 1 – NC KEA Pilot

The first phase of research was conducted on the 2014 NC KEA pilot assessment. It consisted of eight in-depth case studies conducted in schools which closely matched the socioeconomic, racial, and language demographics for the State Board of Education assigned region in which they were located. These case studies consisted of three parts: classroom observations of each participating pilot teacher, interviews with a school and/or district administrator, and a focus group with the pilot teachers. The purpose of the classroom observations was twofold. First, they provided the opportunity for researchers to see the assessment in use. Second, they provided a context to understand feedback from the teacher interviews that may have been unique to particular school/classroom circumstances or characteristics. Researchers conducted the administrator interviews with the school principal and/or a district administrator familiar with the NC KEA. All pilot teachers in each case study school took part in the focus group interviews, and the school's instructional coach was invited to participate if they were familiar with the NC KEA and were supporting its implementation. In total, the eight case studies included: 23 kindergarten teachers, seven school principals, four district administrators, and four instructional coaches.

Researchers visited six of the eight case study schools twice, once within the first 30 days of the pilot and once during the latter 30 days of the pilot. Due to scheduling conflicts, two case study schools were only visited once, and they were visited during the latter 30 days of the pilot. During each school visit, the pilot teachers were observed for no less than one hour of instructional time. Observers first noted classroom characteristics, such as the total number of students present, whether a teaching assistant or co-teacher was present and their contributions to

the classroom, whether the classroom was inclusive of exceptional children, the types of instructional resources and technology available in the classroom and which were utilized during the observation period, etc. After noting classroom characteristics, observers recorded qualitative field notes capturing the teacher's instructional routine, the student's activities, and any NC KEA usage. All observation notes were later transcribed to document format for analysis.

Shortly after the 60-day pilot period, researchers administered an electronic survey that was open to all 305 volunteer NC KEA pilot teachers, administrators, and instructional coaches. The survey provided an avenue for pilot participants outside of the case studies to give feedback regarding the NC KEA and for researchers to triangulate data from the case study interviews and observations in order to identify any potential biases in the smaller case study sample. The survey included 18 closed ended, Likert scale questions and 26 open-ended qualitative questions. The survey questions were directly informed by the case study interview protocols and researcher observations. The survey received 72 total responses, including 52 teachers, 16 administrators, and 4 instructional coaches, with representation from 33 of the 51 participating pilot districts (Ferrara & Lambert, 2015).

All observation field notes, interview transcriptions, and survey responses were uploaded to NVivo 10 so that researchers could perform a grounded discourse analysis on the qualitative data. In this form of analysis, researchers note prominent themes in the data by assigning codes, a word or short phase summarizing a unique idea, to each observation note, interview transcript, and survey response. Three researchers trained in this form of qualitative analysis coded the data independently, after which the codes were compared to ensure inter-rater reliability of the analysis results. This coding process generated 193 unique codes with 3,952 individual references to those codes across all data sources. Some of the codes were categorical (who, what,

where, etc.), while others identified individual topics or themes (training, assessment content, application, etc.). Finally, we conducted frequency and cross-reference/matrix analyses on the coded data to identify the most frequent feedback threads and areas where codes often intersected (Ferrara & Lambert, 2015).

In addition to these qualitative case-study and survey analyses, researchers reviewed the 12,544 pieces of evidence submitted by the pilot teachers to the electronic platform during the 60-day pilot assessment period. Researchers noted the type of evidence (i.e. anecdotal note, photograph, video, student work sample, etc.), which progression the evidence was assigned to, the number of children associated with each evidence, and whether it contained enough specific information for a teacher to accurately assign a learning status based on the evidence. After coding these elements for each piece of evidence, researchers performed descriptive statistical analyses to gain an in-depth picture of how teachers utilized the online assessment platform and teachers' understanding of the assessment process (Baddouh, Lambert, & Ferrara, 2016; Ferrara & Lambert, 2015).

Of the numerous themes identified during these analyses, researchers highlighted four overarching findings in their final technical report on the NC KEA pilot (Ferrara & Lambert, 2015):

1. Practitioner requested improvements to NC KEA specific professional development/training.
2. The need for greater teacher support in identifying quality formative assessment evidence and applying that data to inform their instruction.
3. The importance of district and school administrator buy-in to the utilization of the NC KEA as intended.

4. Practitioners perceived a misalignment of the NC KEA's content, structure, and purpose with other state-mandated curricula and assessments.

Professional Development

While participants walked away from training generally understanding the purpose and content of the KEA (32 interview references from 20 case study teachers, and 63% of survey responses), they felt they were unprepared to use the assessment's electronic platform to upload documentation and enter student learning statuses (22 interview references from 10 case study teachers, and 60% of survey responses). Of those who did feel comfortable uploading evidence to the platform and marking student learning statuses, the majority felt they could not use that data in a meaningful way to inform their instruction (39 interview references from 13 case study teachers, and 57% of survey responses). One teacher summarized this perception quite succinctly: "I'm putting all of this information in, but I'm getting nothing out." This indicated that pilot teachers struggled to move from documenting student knowledge and/or skills to the interpretation and application of that data to inform instruction. This mirrors other formative assessment and KEA studies nationally which found that teachers often struggle to practically apply formative assessment data (more on this in our next report section). Given this practitioner feedback, researchers recommended the following adjustments to future NC KEA specific professional development (PD):

1. A significant portion of PD sessions should be dedicated to hands-on practice with the assessment's online tool.
2. Sample student evidence should also be provided to enhance practitioner understanding of how to upload documentation. This also provides practitioners the opportunity to interpret with peers what learning status should be assigned to

the evidence, enhancing practitioner understanding of the assessment content and inter-rater reliability of status placements.

3. PD should also provide practical examples of how to interpret entered data from a whole-child perspective (i.e. across progressions) and how to then apply that data to teachers' instructional decisions.

Identifying, Interpreting, and Applying NC KEA Data

During the review of evidences entered by the pilot teachers, researchers noted whether each piece of evidence contained specific enough information for a teacher to reasonably make a placement on its assigned progression. For instance, in regard to the crossing midline progression, an anecdotal note stating "Mary danced today" was coded as non-specific, whereas a note stating "Mary crossed midline while dancing the Macarena today" was coded as specific. Researchers found that only 50.2% of the evidences entered contained specific information. This could partially account for why only six of the 23 case study teachers interviewed and 43% of survey respondents felt that NC KEA data was meaningful for driving their instruction. Without strong evidence to interpret, teachers would find it difficult to see the application of that evidence to their instruction. Furthermore, while all case study and survey participants saw value in the assessment's whole-child approach, 57% could not identify ways to use such data for instructional planning. These findings mirror those of other recent scholarship on the formative assessment process that indicates that teachers struggle to move from gathering evidences of learning to using such data to inform their instruction (see review of the literature by Akers, Del Grosso, Atkins-Burnett, Monahan, Boller, Carta, & Wasik, 2015; Little, Cohen-Vogel, Sadler & Merrill, 2020). It is unclear, however, what teachers struggle with specifically in the process of moving from gathering formative assessment data to interpreting and applying it to instruction.

This finding further reinforced researcher recommendations to provide more direct training on the identification of quality evidences of student learning, interpretation of formative assessment data, and finally the application such data as a part of any formal PD provided to support assessment implementation.

Administrator Understanding and Buy-In

When researchers asked participants whether additional, non-KEA specific professional development would be beneficial to support statewide implementation of this formative assessment, an interesting pattern emerged. There were 49 total references describing a need for additional early childhood education training that both outlined and modeled best practice in early childhood instruction. We cross-referenced those statements with the characteristic codes to determine who made these suggestions and whom they identified as needing that additional training. All 49 references came from pilot teachers who identified either their school principal (32 cross-references) or a district administrator (17 cross-references) as needing this additional training. Each reference stated that the administrator did not have early childhood training or teaching experience and, therefore, could neither see the value of the KEA assessment process nor adequately support the teachers during the implementation of a developmentally focused assessment. Due to this lack of understanding among their administrators, a few teachers also voiced concerns that the inclusion of activities to foster social-emotional and physical development in their classrooms would be interpreted as ‘playing’ by their administrators which would negatively affect their performance reviews (10 total references from four case-study teachers and six surveyed teachers). Based on this feedback, researchers recommend that agencies aspiring to implement formative assessment initiatives require at least one district administrator and all elementary school principals take part in a separate professional

development session that provides: 1) foundational knowledge in developmentally appropriate educational practice in early childhood classrooms, 2) an overview of the five domains of early childhood development, and 3) an explanation of how implementing the formative assessment process with fidelity would assist kindergarten teachers in supporting the instructional needs of their students.

Perceived Misalignment with Other Curricula and Mandates

Overall, pilot teachers felt the NC KEA modeled best-practices for the assessment of young children, yet it did not fit well with current state and district mandates which limited their ability to perform the assessment process as intended (39 cross-references of developmentally appropriate/misalignment with other curriculum and assessments from 19 case study teachers and 19 surveyed teachers). Teachers stated that due to tightening accountability guidelines in literacy and mathematics they did not have the time or flexibility to incorporate activities to focus specifically on their students' needs in non-academic areas, such as social-emotional and physical development (39 total references from 18 case study teachers and 19 surveyed teachers). Based on this feedback, researchers recommend that communications during the scale-up to statewide implementation of new formative assessment initiatives explicitly explain how the assessment process fits into the overall educational landscape. In other words, practitioners need to be told directly how the formative assessment supports student achievement on other district- and/or state- mandated curricula and assessments.

Research Phase 2 – Initial Statewide Implementation

The second research phase was conducted in the fall of the 2015-16 academic year, the first year of North Carolina statewide implementation. This study built upon the 2014 pilot research by capturing practitioner perspectives on the changes to the finalized NC KEA content

and process that were informed by the pilot research. As with the previous pilot study, researchers conducted in-depth case studies that included classroom observations, district/school administrator interviews, and focus group interviews with kindergarten teachers and their instructional support staff. Three original pilot case study schools were included in this phase and an additional three schools not involved in the initial pilot were added for a total of six case study schools in three school districts. These schools were selected based on their close representation of the broader North Carolina kindergarten student population in the areas of socioeconomic status, ethnicity, and language demographics. In total, this phase of research included 19 teachers, two district administrators, six school principals, and five instructional coaches.

In contrast to the pilot study, the six case-study schools were only visited once during the latter 30 days of the 60-day assessment window due to scheduling conflicts. All classrooms were observed for no less than one hour of instructional time between the hours of 8:00 a.m. and 11:00 a.m. Observers followed the same protocol as the pilot study in capturing classroom characteristics and then taking qualitative field notes documenting the teacher's instructional routine, the students' activities, and any direct use of the NC KEA. These field notes were later typed into document format for qualitative analysis.

All teacher, school administrator, and district administrator interviews were conducted following the classroom observations and followed a semi-structured interview protocol. The school and district administrator interviews were conducted one-on-one. The teachers were interviewed as a focus group and their instructional coaches were invited to participate if they were involved in supporting NC KEA implementation. All interviews were audio recorded, and then transcribed for later analysis.

Following the 60-day assessment window, researchers conducted an electronic survey that was open to all NC kindergarten teachers. The survey instrument included a total of 22 closed-ended or Likert scale questions and 25 open-ended or free response questions. Some questions were situational and therefore not all teachers were required to answer. For instance, if a teacher responded that they did not participate in the 2014 KEA pilot, the survey system automatically skipped the next three questions which asked teachers to compare their pilot experiences with the finalized NC KEA process implemented in 2015. A majority of the questions were identical to those used in the semi-structured teacher focus group interviews, while others were duplicated from the survey issued during the pilot study to allow for pilot vs. statewide implementation analyses. The survey closed with 736 responses representing 102 of North Carolina's 115 school districts (Ferrara & Lambert, 2016). This survey also sought to gather in-depth teacher perspectives of the assessment outside of the three case study districts by soliciting volunteers for a follow-up telephone interview. Of the survey participants, 106 volunteered to be contacted for a follow-up interview, of which 43 interviews were completed with representation from 26 school districts (Baddouh et al., 2016). These interviews were conducted through Skype and audio recorded for later transcription and analysis.

All observation field notes, interview transcriptions, and survey responses were uploaded to the qualitative data analysis program NVivo. At the beginning of the analysis, researchers imported the codebook generated during the NC KEA pilot research (Ferrara and Lambert, 2015). We then used a grounded approach to analyze the data. In other words, though we partially reused a previously generated codebook, we did not enter into the analysis with particular hypotheses and allowed the data to dictate which codes were used or discarded. Of the 193 codes in the imported codebook, only 104 remained relevant. An additional 48 codes were

generated during this round of analysis, for a total of 152 unique codes used. Some codes were categorical (who, what, where, etc.), while others were indicators of specific topics or ideas (elements of professional development, specific aspects of the KEA content, etc.). There were 35,671 references in total to those unique codes across all data sources. Finally, we conducted frequency and cross-reference/matrix analyses on the coded data to identify the most frequent feedback threads and areas where codes often intersected. Of the numerous themes identified during analysis, researchers highlighted the following findings in their final technical report on the initial statewide implementation of the NC KEA (Ferrara & Lambert, 2016):

1. Professional development provided for the statewide implementation of the NC KEA was highly variable both across and within districts which caused significant issues for implementation fidelity.
2. The utilization of only three construct progressions during initial statewide implementation lowered the value of the assessment in practitioners' minds as it lost the whole-child focus which set this formative assessment apart from other state mandated instruments.
3. Significant changes to the electronic platform are required to make it user friendly enough for teachers to find value in utilizing the tool.

Professional Development

In an effort to limit any undue burden on districts during the initial year of statewide NC KEA implementation, OEL allowed each district to develop their own implementation plans based on their unique capacities. This plan included developing their own professional development models and resources (Ferrara & Lambert, 2016). Given the diversity between districts in regard to their number of kindergarten teachers, the geographical spread of schools,

the time and resources they had available, etc., it is no surprise that the training plans they developed were as individualized as the districts themselves. NC KEA specific training sessions ranged from multi-day workshops that included hands-on time with the assessment platform on either a computer and tablet device (six case study teachers and 42 surveyed teachers), to a single hour set aside during a teacher planning day or weekly planning meeting where the assessment was introduced and the teachers were given the opportunity to ask questions (two case study teachers and 102 surveyed teachers).

A significant number of surveyed teachers (95 teachers or 12.9% of survey respondents) stated that they did not receive any formal training prior to implementation. These teachers received an email from a school or district administrator which told them that the NC KEA was a new mandate for the year and then directed them to either review attached materials (the OEL developed construct progression manual) and/or to follow links to online resources (videos and webinars explaining the electronic platform). These teachers were provided no explanation as to the purpose of the assessment or its utility in informing instruction for their students; therefore, they did not have a proper understanding of how the NC KEA process fit into the overall instructional picture of their district, school, and classroom. Not surprisingly, then, these teachers saw little value in the assessment process and instead felt they were handed, as one teacher describes, “one more thing that added to our already overflowing plate of required assessment and instructional tasks.” Of those 95 teachers, 55 received some form of training after the first month or two of school (i.e. around half-way through the 60-day assessment window). The other 40 teachers never received any instruction regarding the assessment and either struggled through the process by working with their fellow kindergarten teachers or by simply “reading the manual and playing with the program until I learned the ropes.”

Training variability was found not only between districts but also within some districts. Generally, this occurred in larger districts utilizing a train-the-trainer model for each school rather than utilizing a district wide professional development model. For instance, in one large district teachers reported the following: 13 teachers indicated they were lead teachers that attended a half-day training hosted by the district and were then tasked with training their peers, 18 teachers stated they received training during a PLC meeting from their instructional coach or a lead teacher, 12 teachers indicated they had a "brief discussion" that provided an overview of the assessment during a meeting at the end of the last school year, eight teachers received an online video/webinar tutorial which they were instructed to watch on their own time, and 25 teachers stated they received no training at all. Given this widespread training variability and the negative effects that occurred with practitioner understanding and fidelity, researchers recommend that agencies looking to implement formative assessment initiatives on a large scale develop a training course with standardized core elements and materials. Districts should have flexibility to modify the delivery methods as necessary to work within their unique capacities, but the content and resources need to remain consistent to ensure practitioner understanding and fidelity.

Construct Progression Selection for Initial Implementation

While the full NC KEA utilizes a whole-child approach by including progressions within all five domains of early childhood development, initial statewide implementation only required the use of three of these progressions: book orientation, print awareness, and object counting. Several factors influenced the decision to use only those three progressions during the first year. First, the NC legislative mandate governing the NC KEA specifically requires data be gathered regarding children's literacy and mathematics skills at kindergarten entry. Second, feedback

from several 2014 KEA pilot study teachers indicated that utilizing all 10 progressions was overwhelming while they were concurrently learning to use the assessment's electronic platform (Ferrara & Lambert, 2015). Furthermore, a number of pilot study teachers stated that they had an easier time identifying opportunities to elicit evidences of learning regarding literacy and mathematics progressions in the context of normal classroom instruction than for progressions outside of the language and cognitive development domains (Ferrara & Lambert, 2015). Based on that information, OEL chose to implement only three progressions in the first year, and then add the other progressions in phases over subsequent years. The remaining seven progressions were available during implementation. Some districts opted to have all their teachers utilize specific optional progressions, while others allowed their teachers to opt-in or opt-out of using the optional progressions as they saw fit to support the individual needs of their students. This meant that the initial implementation experiences of teachers varied widely in respect to the number of progressions they were required and/or optionally chose to utilize.

The inclusion of only three progressions had unintended consequences on teachers' perceptions of the NC KEA. Teachers overwhelmingly viewed the content of the assessment as redundant and a duplication of data gathered through other state and/or district mandated assessments (12 case study teachers and 357 survey teachers). Further contributing to the teacher perceptions of duplication were issues stemming from the inadequate training many teachers reported they received. For instance, teachers were instructed to use data from these other assessments as evidence for the progressions. Many teachers understood this to mean that these assessments were the only sources of evidence they should use rather than adding observations and student work samples from other instructional activities. Furthermore, only 264 of the 736 teachers surveyed (35.9%) stated they received hands on training on the electronic platform

before or during the 60-day implementation window. This indicates that most teachers did not receive adequate instruction on how to use the reporting functions in the electronic platform, so they did not have the skills necessary to utilize the data they were entering to inform their instruction. These issues converged in such a way that many teachers interpreted the KEA to be no more than a database to "house multiple sources of assessment data for the state's use," rather than a formative assessment process with an electronic platform that could assist them in individualizing instruction for their students. Furthermore, the academic focus of the three progressions selected for initial implementation negatively affected teachers' perceptions of the value of the assessment process, as the selection ran counterintuitive to the message provided during professional development and in formal communications about the NC KEA that described it as whole-child focused. Based on these findings researchers recommend that agencies which opt to introduce formative assessment content in phases select a balance of academic and developmental domains to ensure a continued focus on whole-child instruction. Additionally, professional development should teach practitioners how to interpret the assessment data holistically (i.e. across domains, not just within) via instruction on the reporting features of the electronic platform.

Electronic Platform

More than half (53.3%) of teachers interviewed and surveyed indicated that the NC KEA website was either difficult or very difficult to use. The most common feedback in explanation of this perception was the sheer amount of time it took to upload evidences to the website and assign student learning statuses (12 case study teachers and 65 surveyed teachers). Teachers stated that there were "too many clicks" and they had to continually go "back and forth between multiple tabs or pages in order to complete the uploading and finalization process." Furthermore,

teachers had difficulty accessing the website from their school computers due to compatibility issues with school internet security software or protocols and/or the website's integration with NC PowerSchool (six case study teachers and 66 surveyed teachers). A total of 20 teachers from multiple districts indicated that they did not gain access to the system until the final week of the 60-day implementation window, so the task of entering all of their evidence was both overwhelming and useless to driving their instruction since the data was not current. Proportionately more teachers found the NC KEA tablet and smartphone app to be user friendly, with only 24.6% of the surveyed teachers stating it was difficult or very difficult to use. The most common teacher feedback regarding the iOS app was that they did not receive enough training on how to use it, but it became easier to use once they "played around with it" (33 surveyed teachers). Based on this feedback, researchers recommend that agencies opting to utilize an electronic platform or application as an assessment tool provide enough lead time for districts to ensure their technology infrastructure can support the new software, install it on necessary devices, and troubleshoot connectivity issues prior to initial implementation. Furthermore, the software must be user-friendly enough for teachers to view it as a helpful tool and not "just one more thing added to [their] overly full plate that eats up valuable instructional time."

Research Phase 3 – Implementation Teams, Feedback Loops, and Structures of Support

From the assessment's inception in 2014, OEL strategically built their implementation plan on the foundations of implementation science most notably outlined by Dean Fixsen and his research team at the National Implementation Science Network (NIRN). This implementation approach called for the development of a multi-tiered teaming structure that allowed open communication and periodic feedback to flow from the state down to the building level and vice

versa (Fixsen, Blase, Metz, & Van Dyke 2013; Fixsen, Blase & Wallace, 2009). OEL expended significant effort and resources to promote the development and capacities of these implementation teams, however, there was little understanding of practitioner perceptions of their utility in supporting NC KEA implementation. With that in mind, the third phase of CEME's NC KEA research shifted focus from classroom level implementation to researching the implementation team structures put into place at the state-, regional-, district-, and building-level.

Researchers again utilized a case study approach for data collection. Four North Carolina school districts were selected as case study locations based on a number of criteria, including: their State Board of Education (SBE) region affiliation, the socioeconomics of the district's kindergarten population, the urban-centric locale for the district, and whether the district took part in a usability study for a potential scale-up of the formative assessment process to 1st through 3rd grade. Usability study participation was an important factor to consider during district selection, because usability districts received additional support developing and sustaining their implementation teams through monthly meetings led by OEL staff. In order to determine how this additional support affected district and building teams, two usability districts and two non-usability districts were included in the case studies.

During data collection, researchers divided into two groups. The first performed implementation team meeting observations and member interviews at the state and regional levels. The second performed the same tasks at the district and building levels. Meeting observers recorded qualitative fieldnotes, and interviewers utilized a semi-structured interview protocol for all interviews. Meeting observations afforded researchers the opportunity to record the unique team makeup, dynamics, and procedures for each implementation team and provided context to understand feedback provided during the individual team member interviews. In total,

researchers conducted eight implementation team meeting observations: one state implementation design team (SIDT) meeting, three regional implementation team (RIT) meetings, and four district implementation team (DIT) meetings. Building level implementation teams were not yet established in any of the sample districts, therefore there were no observations at the building level. Interviews were conducted with 30 individual implementation team members, five SIDT members, 16 RIT members, four DIT members, and five building level implementation leaders. Regional implementation team interviews were held with the two OEL regional consultants leading each team, as well as two additional members from each RIT. To avoid the potential for overrepresentation at the regional level during analysis, each of these interview pairs were aggregated and treated as a single interview. Researchers used grounded theory in developing a codebook for data analysis, meaning that the data itself drove the analysis rather than entering with specific hypotheses to test against the data collected. Prominent, repeated themes were created into codes, then a Yes/No method was adopted to identify whether that theme was present in each observation and interview. Finally, a percentage was calculated by implementation level to see to what extent the coded theme was present and discussed. This method of analysis allowed for equivalently weighted comparisons across implementation levels regardless of the disparity in interview numbers between them. Of the numerous themes identified during analysis, the following findings were highlighted in the final technical report on this implementation team research (Ferrara, Lin, & Lambert, 2017):

1. An apparent disconnect exists between the state/regional teams and the district/building teams in their view of the NC KEA as aligning with other state/district mandated curriculum and assessments.

2. A lack of administrative understanding and buy-in for the NC KEA continues to be a barrier for implementation.
3. All implementation levels view professional development as a means to overcome implementation barriers; however, a difference in the conceptualization of what constitutes professional development led to a difference between state/regional teams and district/building teams in their preferred frequency of lessons/training.
4. District/building teams continue to request clarity around NC DPI's vision for early education, and an explicit explanation about how the NC KEA fits in with other state/district mandates and standards.

NC KEA Alignment and Integration

While state and regional level practitioners view the NC KEA as aligned with the demands of other state mandated assessments in early childhood classrooms, practitioners at the district and building level disagreed. In fact, district and building level implementers overwhelmingly viewed the NC KEA as not only misaligned, but often conflicting with the requirements and demands of other required assessment tasks. For example, one regional level participant commented, "People said they feel like [another state mandated assessment] is at odds with our assessment – ours is developmentally appropriate and is really what kindergarten children should be doing, but the other is required, and monitored. Because that assessment has 'teeth' in terms of being tied to teacher's evaluations, they put more emphasis on it." This perceived misalignment is a repeated theme from the previous two phases of research. This feedback highlights a need for agencies attempting to implement developmentally focused formative assessments into the current academically focused accountability climate to provide an

explicit understanding of how the initiative not only fills a potential instructional or assessment gap, but also helps support student achievement on other mandated measures.

Administrative Buy-In

Practitioners at all levels reported that a lack of administrative buy-in for the NC KEA acted as a significant barrier to implementation. Several factors likely contributed to this perceived lack of administrative buy-in. First, practitioners noted that many of the administrators tasked with overseeing NC KEA implementation in their district or school lacked training and/or experience in early childhood education. One instructional coach noted "the building administrators that we have come from a place where understanding is a little limited. They have, at the district level, realized that in order to get the change needed from our principals and the acceptance of different practices in the classroom, we've got to educate the administrators." This lack of exposure to early childhood educational foundations limited administrators' ability to understand the purpose and value of the NC KEA, and by extension limited their ability to support their instructional staff's implementation efforts. Secondly, the pressures placed on administrators to meet state required end-of-year accountability goals influenced their priorities. As mentioned before, many district and school level practitioners viewed the NC KEA as misaligned with other mandated assessments, so administrators dedicated greater resources and time to support what they felt were higher priority instruction and assessment tasks. As one instructional coach noted, "If there is not support from the building administration, then it's not perceived as a priority at the administrative level, and the teachers are not going to have it as a priority either." These priorities could also explain why a shortage of resources to support NC KEA activities is noted much more frequently by district and school level practitioners than by those at the state and regional levels. This feedback mirrors findings regarding a lack of

administrative buy-in and its negative effects on implementation efforts from the previous two phases of NC KEA research. This indicates that agencies attempting to implement new formative assessment initiatives should prioritize professional development and messaging efforts for administrators early in the implementation process in order to garner their support and smooth the way for their instructional staff to implement the process as intended.

Professional Development as a Means to Overcome Implementation Barriers

Practitioners at all levels agreed that continued professional development should occur with some regularity to ensure the success and sustainability of the NC KEA; however, their opinions differed in how often it should be offered. State and regional practitioners preferred PD to happen on an ongoing, daily, or weekly basis, while district and school level practitioners preferred PD sessions to occur every few weeks, to every few months. This difference in preferred PD frequency is potentially due to a difference in the conceptualization of what constitutes professional development. State and regional practitioners referred to PD in a more general sense, including discussing NC KEA related activities in coaching/mentoring sessions and in grade level or PLC planning meetings:

> "Ideally, a district would have a team that developed a comprehensive professional development plan that includes both training and coaching. The training would include practices that have research-supported high outcomes; training methods that yield a better chance of the practice being used by the teacher. The district would also gather information about how effective the training/coaching is so they can improve upon that to better support teachers."

District and school level practitioners, however, defined professional development as a separate event of some type, whether a short meeting dedicated specifically to discussing the NC KEA or

a half- or full-day training workshop. This could account for why a majority of state and regional practitioners suggested a coaching model as a preferred PD delivery method (SIDT 50%, RIT 80%), while no district and school practitioners mentioned that method. District and school practitioners simply may not consider mentoring/coaching as professional development since it is relatively informal. District and school practitioners who work closely with curriculum specialists and instructional coaches often mentioned how invaluable they were in supporting implementation (District 50%, School 100%). This suggests that district and school practitioners may be open to the coaching model the state and regional teams mentioned. With that in mind, researchers recommend utilizing a coaching model to support PD efforts. Furthermore, a coaching model can help support implementation fidelity by incorporating the use of fidelity measures designed to identify areas of the formative assessment process where individual practitioners may need additional guidance, training, or support during implementation.

Clarity of North Carolina's Vision for Education in the Early Grades

District and school practitioners often requested some type of roadmap or guide to better understand how the NC KEA fits into the state's vision for elementary education. This request is likely due to the persistent perceived misalignment of the NC KEA with other state/district mandated assessments. The lack of clarity around instructional priorities for early childhood classrooms from NC DPI must be reconciled for practitioners in order for them to both comprehend and buy-into the value of this formative assessment process. This issue is clearly illustrated in the following interview exchange with an administrator tasked with directing NC KEA implementation in her district:

> District Administrator: "I think there is this sort of ambiguity about what our focus needs to be. Is it whole child and developmental, or is it all about the academics? When we

have two seemingly different asks, different pressures and priorities, which is most important and how do we balance that?"

Interviewer: "How do you address that ambiguity, that haziness they have about these 'seemingly different asks?' It seems like you're providing good information, but still some dots aren't connecting for them."

District Administrator: "I think we aren't connecting them well, because they aren't connected well for us. Frankly, I have these questions myself."

Without a unifying vision, administrators may continue to prioritize time, resources, and support for other assessments and initiatives they view as more important to the overall performance of their school on annual accountability measures. A strong, clear message from the state passed down through the communication frameworks already in place between the implementation team levels, could go a long way in assisting regional, district, and school level implementers facing administrative push-back and lack of buy-in in their respective areas.

III. Previous Research on Formative Assessment and KEAs Across the United States

Formative assessment is a process requiring teachers to recognize and interpret demonstrated skills, followed by providing continuous scaffolding and feedback to students (Heritage, 2013). According to Shepard (2009), formative assessment is of little use if teachers do not know what to do when students are unable to grasp important concepts. Research demonstrated that teachers are much better at identifying what is being assessed and the level of student understanding than they are at determining appropriate next steps for children. Teachers need more than good assessment instruments, they need help to develop methods to interpret and act on results in a formative way (Shepard, 2009). It is beneficial to provide resources to help teachers learn formative assessment techniques, including using information to intervene with students who do not yet understand key concepts.

Teachers are continuously adapting to new expectations to implement new assessments throughout their career. While teachers are frequently administering new forms of assessments, ideally teachers are making best efforts to use assessments to gain more understanding of each individual student he/she instructs (Schachter, Strang, & Piasta, 2019). Information from this assessment guides instruction and helps teachers ascertain the level of impact instruction could be having on their class. All of this holds true for NC kindergarten teachers as well when conducting the KEA.

In reality there could be teachers simply going through the motions when delivering statewide assessments. It is not uncommon for wide-scale assessments to occur without teachers and schools considering how or why data from these assessments can serve the purpose of benefiting teachers and students (Lesaux & Marietta, 2012; Ohle & Harvey, 2019; Young & Kim, 2010). In one study, less than half of education professionals implementing NC KEA

discussed using data from the assessment for instructional planning (Ferrara et al., 2017). Another study focused on NC KEA implementation and data usage as reported by NC principals and kindergarten teachers that some elementary schools in the study reported not using data from the assessment at all beyond mandated implementation requirements (Little et al., 2020). Considering this lack of data usage, it is possible that teachers and administrators have misunderstood the purpose of the NC KEA and therefore they do not view the data the process provides as valuable to informing instruction for their students (Ferrara & Lambert, 2015).

Assessments adopted with the intention of improving student learning and a teacher's experience are critical aspects of education. However, when teachers do not accept or view an assessment as worthwhile it leads to increased stress (Lambert, 2019). On the other hand, when teachers view an assessment as useful it can alleviate stress. While it is critical for teachers to support a widely used assessment, it is important for instructional support staff (i.e. teacher coaches, facilitators, curriculum specialists) and school and district level administrators to buy in as well.

Teachers need a full understanding of learning progressions and classroom experience observing what constitutes a student demonstrating understanding (Heritage, 2013). This is a skill that is built over time and can be developed with the assistance of a more experienced educator. Instructional support staff members provide assessment expertise and assist teachers in utilizing information from formative assessments. Interactions with more experienced educators creates professional accountability for teachers to adjust instructional practices (Young & Kim, 2010).

A group of surveyed teachers piloting the NC KEA in 2014-15 suggested that school and district level administrators needed additional training on this formative assessment process

(Ferrara & Lambert, 2015). It is possible this is because many administrators come from upper elementary or secondary teaching backgrounds. More specifically, providing administrators with training focused on providing kindergarten teachers support during assessment implementation. It is important for administrators to understand assessments across all grade levels in schools they serve in order to offer support for teachers and ensure assessments are implemented as intended.

Additionally, administrators serve as models for how to use data and often provide validation to teachers in how they are gathering and interpreting data (Harvey & Ohle, 2018). Administrators have the opportunity to provide consistency in expectations for gathering and using data from assessments. Young (2006), encouraged district level leaders to take initiative in building a school leader's ability to implement changes leading to an improved data usage climate across a school. Research showed when school administration did not have early childhood training or teaching experience in early childhood grades, they were less likely to see the value of the assessment and were not prepared to adequately support teachers during implementation. Harvey and Ohle (2018) offered the suggestion for principals to use frequent check-ins with teachers in the form of simply asking "how can I help you with this? how is this data informing your instruction? or what did you learn about your students that we should know about?" (p. 20) when interpreting assessment data. These questions provide opportunities for teachers and administrators to build comfort with the assessment, reflect on how data gathered from the assessment is impacting instruction, and opening avenues for collaboration with other education professionals.

In a 2006 study conducted by Coburn and Talbert, researchers found district administrators more interested in the overall psychometric properties of assessments and if an

assessment demonstrated desired outcomes. In contrast, researchers determined teachers and administrators working directly with teachers are more influenced by insights into student thinking and reasoning and teacher judgment as sources of validity. These conceptualizations provide a beginning framework for looking at how uses of data differ by organizational level in school districts. Pressure from accountability goals tied to end-of-grade assessment performance in third through fifth grades impacted the prioritization of the NC KEA (Ferrara et al., 2017). Implementation teams with strong leaders and relationships across teams allowed for effective implementation of the NC KEA and use of data to target instruction.

In order to achieve the benefits of data-use, teachers must find data useful for instruction. Teachers need more clarity regarding the content of readiness assessments and how assessment content links to learning standards. This link demonstrates how data fits into the curriculum and learning goals of students (Schachter et al., 2019). Having a better understanding of the assessment may help teachers interpret how the content relates to overall learning goals. Schachter et al. (2019) reported that when teachers have time and support in understanding data they build successful data use practices. According to Young and Kim (2010), no matter the strategies designed to prepare teachers and schools to use assessment data, the data "must have legitimacy with teachers" (p. 18).

School and district level organizational factors have the capacity to influence the interpretation and usage of data from formative assessments. District and school administrators play vital roles in setting expectations for data usage across schools. Even further into the school-level, grade level norms impact the way teachers interpret and prioritize using data. As with many educational practices, leadership influences loosen or tighten how teachers interpret and use data (Young, 2006).

The existence of school and district influences points towards a need to involve all stakeholders in understanding the purpose and intent of a widely implemented assessment. Ohle & Harvey (2019) noted the use of KEAs should be evident across these two levels. At the school level, KEAs offer many opportunities to collaborate through supporting teachers, overall program planning, individualizing instruction for student success, and allow for teamwork and partnerships to form amongst colleagues. At the district level, KEAs allow leaders to notice readiness patterns that will help decision-makers allocate resources and support systems across schools as needed.

The amount of support a teacher needs may vary based on total years of experience, years of experience within the grade level, and years using the assessment. Leaders of schools in the developmental stages of using data can structure team interactions with instructionally relevant activities so that teachers can practice data analysis while simultaneously forging new collaborative norms (Young, 2006). Data driven decision making is a continued area to target for improvement, with a focus on data utilization and vertical data sharing (Little, Cohen-Vogel, Sadler, & Merrill, 2019).

Teachers need support in developing a capacity for data use. Educators feel unprepared to use data to engage in activities like adjusting the curriculum and interrogating data in meaningful ways. Schools can have their own norms in regard to data usage, some may be more open to data sharing and professional discussion around data sharing. Teachers and school administrators may interpret assessment data in different ways, when this occurs the principal's perspective usually prevails as the "correct" interpretation (Little et al., 2020). Formative assessment measures can help teachers get to know children at the beginning of the academic year. These allow for

teachers to understand the strengths each child brings to the classroom, provide feedback to children, and identify areas where each child needs support.

As mentioned earlier, it is important for instructional support staff and administrators to actively support teachers in implementing and understanding data from the KEA. Roehrig, Duggar, Moats, Glover, & Mincey (2008) explained the teacher coach's role in helping teachers make sense of data. The coach helped teachers access and interpret data and make informed links to curricula. Barriers to the practice of using data included lack of time and classroom management difficulties. Even with professional development, a disconnection can occur for many teachers between the time they complete training and when they shut their classroom door and attempt to apply what they have learned.

The goal of providing data to teachers is to help them identify and adjust instruction for students who are demonstrating difficulties. Knowing students' skills is key for selecting and implementing effective instruction. Research on effective teachers and effective schools converge on the notion that student assessment data are frequently gathered and considered by teachers and schools with the best student outcomes (Roehrig, et al., 2008). In schools considered successful in educating all students, regardless of race or class, education professionals analyze the results of assessments in teams to improve teaching and learning.

Roehrig et al. (2008) reported teachers discussed monitoring student progress and areas of strengths and weaknesses, adjusting or forming groups for individualizing instruction, and identifying appropriate activities, intensity, and level of instruction. Teacher knowledge and coach availability can affect the successful implementation of the phenomenon. Principals are critical in developing and sustaining data usage cultures by supporting coaches and teachers in the use of data to inform instruction.

Conclusion

Assessment practices in kindergarten, such as with the implementation of the NC KEA, help address the challenge teachers face when receiving little to no information about incoming students. Supporting educators in assessing student abilities is essential so that teachers can support learning and development. Whether and how teachers use data to guide instruction is dependent upon assessment practices from the individual classroom up to the state level. There are many connected factors worth considering when implementing a formative assessment. Teachers need to understand the purpose of the assessment while also understanding the progressions and the overall content addressed within the assessment. While the classroom teacher is the key actor in successful implementation and adoption, many other education professionals play vital roles in facilitating the formative assessment process.

IV. Preparing Teachers to Make Valid Placements on Developmental Progressions

Assessment measures that yield valid information are essential instructional resources for teachers. Valid placements on developmental progressions can be very useful for instructional planning and supporting the growth and development of young children. When the formative assessment process (FAP) is working well in classrooms that serve young children, teachers can focus on what is important for the developmental progress of each child. However, teachers need to be equipped with a specific set of skills to make valid placements on developmental progressions. Furthermore, teachers will most fully engage in the FAP, and develop these skills most efficiently, when they perceive the FAP as a resource that can provide meaningful and *useful* information, and thereby can help make teaching easier, more effective, and less stressful.

So what constitutes *usefulness* to teachers of young children? How can teachers of young children use the FAP to gain a valid understanding of the current status of a child's growth and development? How can they know what developmental tasks come next for each child? How can they structure learning experiences that support each child? The answers to all of these questions are embedded within the skills necessary to make valid placements on developmental progressions.

Ideally, the educational process for all children is planned with the ends in mind. As educators, we plan and implement instructional strategies to ensure that all children can realize age-appropriate educational outcomes. We use the information provided by assessment methods to inform the process all along the way. Our instructional objectives are our ends, our pedagogical strategies are the means, and our assessment methods allow us to document progress along the way. Shepard (2000) has argued that the instructional process is incomplete and unstable when it does not include assessment methods that are fully integrated into everyday

teaching practice. To the extent that this assertion is true, formative assessment skills are essential components of quality teaching. Once teachers have mastered a complete understanding of curricular objectives, and have developed at least a baseline working mastery of classroom management strategies, instructional leadership techniques, and effective instructional strategies, they are ready for assessment strategies that can help them understand the children they serve.

However, assessment measures can only fully contribute to the instructional process when they yield valid information. How can teachers know the progress each child is making toward their goals? How can they know what milestones are coming next for each child? How can they be sure they are helping each child have meaningful experiences? How can they ensure that each child maximizes his or her opportunities as they approach the next developmental tasks? The FAP, when implemented with fidelity, can yield valid information to address all of these questions.

Understanding the Purpose of Authentic Formative Assessments for Young Children

Teachers simply cannot implement AFA measures with fidelity without first fully understanding the formative assessment process and the purposes for which AFA measures were developed. Ensuring that teachers have a complete understanding of the purpose of AFA is the first, and perhaps most important step in preparing them to use the measures. The validity of the information that any measure provides is inseparable from the purpose for which a measure was designed. Therefore, it is essential to outline and understand the purposes of authentic formative assessment (AFA) measures in order to help teachers become fully equipped to implement them with fidelity.

Let us then outline the purposes of AFA as a means of identifying the important concepts that teachers must grasp in order to implement the FAP with fidelity. Heritage (2010) defined

formative assessment "…as assessment *for* learning, not assessment *of* learning…" Formative assessment focuses on the learning process and is used to support learning while learning is taking place. Formative assessment consists of the formal and informal processes teachers and students use to gather evidence for the purpose of supporting and improving learning. Formative assessment has been defined as "…a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to help students improve their achievement of intended instructional outcomes…" (AERA/APA/NCME, 2014; McManus, 2008). Black and William (1998) conducted an extensive survey of the research on formative assessment and demonstrated that student learning increases when teachers use formative assessments appropriately. All of these concepts apply to the instructional process across the lifespan from early childhood to adulthood.

High quality early childhood teachers desire to learn as much as possible about their students. For them, the formative assessment process is a natural, ongoing component of daily instruction (NC Construct progressions and situations, NC Office of Early Learning, 2015). Furthermore, developmentally appropriate formative assessment has been identified as a particularly vital component of high quality early childhood education (NAEYC & NAECS/SDE, 2003). AFA measures have been designed to help teachers get to know children at the beginning of the academic year. They help teachers understand the strengths that each child brings to the classroom. They help teachers provide valuable feedback to children, understand their interests, and identify the specific areas where each child needs support (Heritage, 2010). Therefore, AFA measures help teachers contribute to the growth, learning, and development of the children in their classrooms at every point during the academic year (Bredekamp, 2011; Clements & Sarama, 2009). A rich understanding of each child helps high quality teachers plan

classroom activities, select and rotate classroom materials, and individualize and differentiate instruction.

However, without a full understanding of FAP, teachers often operate naturally from working assumptions and evolving narratives about each child in their classroom. When the formative assessment process is operating as intended, teachers can challenge and inform these assumptions and narratives with useful evidence, and thereby transcend the subjective "halos" and "horns" they can too easily place over children in the absence of valid evidence. High quality AFA measures can offer useful developmental progressions to teachers that provide structure and guidance as they attempt to understand and support the unique needs of each child.

AFA measures are designed to yield useful information while embedded within the ongoing work of teachers. As teachers develop learning opportunities for children, they can collect evidences of child progress and gather ongoing portfolios throughout the academic year. In contrast to direct summative assessments, or any artificial testing situations that are designed to elicit specific behaviors, these evidences are collected within regular classroom activities in their natural context. High quality AFA measures are designed to help teachers organize evidences gathered through observations, conversations with children and families, samples of children's work, photos, video clips, recordings, etc. AFA measures are designed to assist teachers as they reflect upon, analyze, and summarize evidences, and identify the best placements for each child across a series of developmental progressions.

Therefore, AFA measures, when fully implemented, help teachers understand how each child is currently functioning based on behaviors observed in the course of naturally occurring classroom activities. They can provide teachers with actionable information that helps identify each child's strengths and areas where more support is needed (McMillan, 2014). They can help

teachers monitor child growth, guide instructional planning, and help teachers set meaningful learning targets for each child (Lambert, Kim, & Burts, 2013). In this way, AFA measures are designed to be a central component of the instructional process and can enhance child growth by helping teachers present children with opportunities to maximize their learning potential (Shepard, 2000).

The information that AFA measures provide can also facilitate communication with children, families, and other stakeholders. High quality formative assessments can even empower young children by helping them feel included and valued in the educational process. For example, the process of evidence formation and collection can include child choice, and thereby engage young children in meaningful dialogue with teachers about their developmental progress (Heritage, 2010). The FAP helps teachers provide high quality feedback to children and thereby enhances the teacher-child relationship. Teacher-child interactions can be much more meaningful and positive when teachers learn to use valid assessment data to target their instructional strategies to the specific needs of each child.

Information from AFA measures can help teachers communicate with parents in terms that are easily accessed and understood. Teachers can point parents to placements on the developmental progressions, and their associated child work samples and anecdotes, which can address child progress with specific examples of what children know and can do. This process can help facilitate rich conversations about the child's development within the classroom, family, and cultural context. Teachers can even solicit evidence of child progress and development from parents and other caregivers to inform the assessment process and ensure that progress is measured accurately. This process can help teachers partner with parents to support the growth and development of each child.

Formative assessment information is also particularly helpful for teachers when they communicate and collaborate with other educational professionals within their professional learning communities. Data and evidence driven conversations can lead to richer interactions with everyone connected to the children. A rich and detailed picture of a child's current learning status and patterns of growth and development can help other educational professionals provide individualized and informed support to the child. Teachers can use these richer conversations to solicit the participation of involved professionals in the evidence gathering process and can gather additional understanding of each child as they seek specific input from educational professionals about how to support children.

This information can also be useful to those who support the professional development of teachers. It can provide an enhanced picture of how a teacher experiences and is aware of classroom processes, observes children in the classroom, and collects evidences of child progress and development. Formative assessment data can provide process information regarding how teachers analyze evidences of child progress, make placements on developmental progressions, and use that information to support child growth, learning, and development. This same data can be very useful as mentors and other support personnel help teachers plan individual, small group, and whole group instructional activities.

Mentors, coaches, and technical assistance providers can also use formative assessment data as a catalyst for rich conversations with teachers that can help them reflect about instructional practice and set professional development goals. This process can help teachers increase their observational skills and become much more aware of how each child learns and functions in the classroom. In this way, mentors can provide data driven support to teachers and thereby model for teachers the process of using data to individualize support for children.

In summary, AFA measures can help teachers reflect about their work, and provide structure to the interpretations they make while observing children in the classroom and can do so in useful, actionable, and practical terms. AFA measures are designed to be helpful resources for teachers that inform the full range of instructional activities across the academic year. From getting to know children at the beginning of the year, to planning instructional strategies throughout the academic year, to collecting evidence that children have gained the skills and abilities that are the goals of instruction by the end of the year, AFA measures can provide useful information to teachers when they are implemented with fidelity.

Specific Steps for Preparing Teachers to Make Valid Placements Using AFA Measures

Messick (1995) contended "…What ***is*** needed is a compelling argument that the available evidence justifies the test score interpretation and use…" He argued that construct validity, therefore, consists of the evidence that can support the "…appropriateness, meaningfulness, and usefulness of score interpretations." He therefore characterized the use of the information that assessments can provide as a process of interpreting scores and using them to make inferences in the context of particular functions, actions, decisions, and purposes.

So what skills do teachers need to make valid placements on the progressions of AFA measures, and use those placements to make valid inferences about the developmental progress of young children? First, they need clear messaging from those requiring them to implement an assessment system. This messaging must communicate the true nature and purpose of the FAP, as outlined above, and must make clear distinctions between AFA and direct summative assessment. Without such guidance, teachers can be very quick to assume that all assessment is direct summative assessment that ***will be used for accountability purposes***. When teachers make

such false assumptions, they effectively and systematically invalidate the information that AFA measures provide.

Next, teachers need to be trained to fully understand the FAP response process. Unlike direct summative assessments, AFA measures have no standardized test administration process. Rather, they rely on a complex response process that involves an interaction between the teacher, the classroom environment, and child behaviors. This process has been described as a continuous cycle of activities that is part of everyday instructional activity in the classroom. This cycle is often outlined in the following phases: 1.) understanding what is next for a child and set learning goals, 2.) defining and understanding criteria that will indicate progress toward the next level of development, 3.) gathering evidences of growth, development and learning, 4.) analysis and interpretation of evidences, 5.) making placements on developmental progressions, and 6.) adapting instruction to support the unique needs of the individual child (Heritage, 2013).

After a child masters a particular level on a developmental progression, then this cycle can repeat itself as the child moves toward the next developmental level on a specific progression related to an instructional objective. This process is also simultaneously playing out over many developmental progressions across a variety of learning objectives and developmental domains. This cycle begins with a data-driven sense of where a child is currently functioning relative to a particular developmental pathway, and progresses through to data-driven support for the growth, learning, and development of the child. It is an integral part of the instructional process and is neither distinct from nor supplemental to learning. Rather, it is the natural manifestation of high quality instructional practices.

Within this complex cycle, the child response process resides in the evidences elicited by the teacher, classroom activities throughout the entire instructional process, and the classroom

environment itself. Therefore, full implementation of AFA measures assumes a minimum quality for the entire instructional process. An under-stimulating, ineffective instructional environment may not elicit as many construct-relevant evidences and behaviors from children as might emerge in a higher quality classroom environment. Children may not demonstrate as much about what they know and can do in lower quality classrooms. Construct-irrelevant variance in the assessment scores yielded by AFA measures can, therefore, relate to almost any aspect of teacher behavior and classroom quality.

It is not only the richness of the child evidences elicited that can vary from classroom to classroom and teacher to teacher, but also the teacher's ability to recognize, select, record, and analyze those evidences. If a teacher misunderstands or misapplies any of the steps in the assessment cycle, they are more likely to be introducing construct-irrelevant variance into the assessment scores. For example, teachers can effectively alter the difficulty of an item or progression for the children in their classroom when they select evidences or artifacts that they consider to be appropriate to support particular ratings on a progression. Their own leniency or strictness can enter the process as they identify what they perceive to be construct-relevant evidences, and match those evidences to the behavioral anchors and indicators that are integral to each developmental progression.

Therefore, the AFA response process will only include evidences that are representative of a child's true abilities when a teacher has mastered the complex set of tasks involved in all phases of the assessment cycle. Teachers have to understand fully how each progression corresponds to learning objectives in the applicable curriculum model and child learning standards that govern their work. They have to understand how to recognize valid evidences that relate to the indicators on the progressions, match those evidences accurately with the

appropriate levels on the progressions, and determine when they have sufficient evidence to support placements on the progressions.

Black and William (1998) underscored this point as they summarized, through meta-analysis, the positive benefits of formative assessment on child learning outcomes, and also noted several issues that can impede the practice of effective assessment for learning. They organized their findings by various components of teacher functioning that relate to the formative assessment cycle. In each of the following areas: choice of classroom tasks, discourse with and feedback to children, open ended questions for children, and choice of instructional strategies in response to assessment information, they noted both positive benefits for children from high quality practices and sources of between-teacher variability that can compromise the validity of the AFA process. These findings echo those of many other researchers who have demonstrated the value of informed feedback to all involved in the educational process (Hattie & Timberley, 2007).

The child response process for AFA measures is so much more complex than it is for direct summative assessments. The FAP assessment cycle requires teachers to master a complex web of behaviors in order to fully implement the AFA process. If they implement any of these tasks with partial fidelity, there is a risk of introducing construct-irrelevant variance into the resulting scores. So what is required to bring teachers to full implementation of an AFA measure so that they can produce valid scores and inferences? In short, effective and accurate messaging needs to be followed by thorough, effective, and intensive training and coaching to support teachers. Content rich training, data-driven coaching, guidelines and safeguards regarding procedural fidelity, and quality standards need to be put into place in each of the following areas:

1. Foundational Principles of Formative Assessment

2. The Content of the Developmental Progressions
3. Connections Between the Progressions and the Applicable Standards for Child Learning
4. The Role of AFA in the Instructional Process
5. Collecting Valid Evidences
6. Collecting Sufficient Evidences
7. Analyzing Evidences and Making Valid Placements on the Progressions
8. Using Electronic Systems to Store Evidences and Placements

First and foremost, those who train teachers to implement an AFA process must begin at the beginning. The first task is to provide training regarding the nature and purpose of the formative assessment process. The messaging that reaches down to the classroom teacher about the purpose and use of AFA measures plays a large role in a teacher's ability to implement AFA measures with fidelity, and thereby produce valid assessment information. Messaging from district and school leadership, along with teacher understanding of that messaging has been shown to be strongly associated with implementation fidelity.

Teacher training about the FAP often focuses on procedural or mechanical issues such as how to use an online electronic portfolio system for secure storage of evidences, and how to enter placements on the progressions into an online data collection system. These features of an AFA system can be helpful resources for teachers, but training in their use does very little to support the validity of the information that an AFA measure provides. A heavy focus on these practical concerns during training communicates to teachers that the FAP is an exercise in compliance with a mandate. Rather, beginning at the beginning requires teachers to gain a full and complete understanding of AFA before they collect any evidences or make any ratings. This requires teachers to understand the many facets of a complete assessment system and how

purposes, functions, appropriate inferences and uses vary across assessment types. Many teachers tend to believe that all assessment ***is direct summative assessment***, rendering the two concepts indistinguishable (Shepard, 2000). Without an understanding of the very particular concept of direct summative assessment, and how it differs from other types of assessment, many teachers will create direct assessment tasks in their classrooms as their central means of collecting evidences for AFA measures (Ferrara & Lambert, 2016). This of course invalidates the ***authentic*** aspect of AFA.

Let us examine the essential messaging for teachers, the necessary precondition for building a sufficient knowledge base to support the implementation of AFA. First, teachers must understand that AFA measures ***are not*** developmental screeners. AFA measures are typically not designed to provide cut scores that indicate the need for further testing or diagnostic processes. Similarly, they have not been designed to lead to specific decisions using cut scores that result in high correct classification or high false positive rates. Furthermore, formative assessment information alone ***is not*** appropriate for making high stakes placements or diagnostic classifications of children and no such decisions should be based on single sources of information. However, the information provided by high quality use of well-developed and validated formative assessments can make valuable contributions to multiple source, multifaceted, multidimensional and multidisciplinary professional discussions of the needs of individual children.

As useful as formative assessment information and processes can be to teachers, they must understand that formative assessment ***is not*** summative assessment. It ***is not*** appropriate to use the information that AFA measures provide about specific children, or groups of children, for any summative purposes such as performance evaluation of teachers, program evaluation, or

assessment of classroom, center, or program quality. It is also inappropriate to use the information yielded by formative assessments to make any kind of high stakes decisions. In fact, attempting to do so can introduce construct-irrelevant variance, and can give teachers perverse incentives to make less than valid placements on the developmental progressions and can thereby rob them and the families and children they serve of the benefits of the appropriate uses of formative assessment information.

Teachers need to understand that AFA measures ***are*** designed to be developmental assessments, meaning that they include progressions of growth, development, and learning that describe a sequence of stages and behavioral anchors that children are generally expected to demonstrate. Each progression includes these descriptive anchors that illustrate behaviors, work samples, and other evidences that can be observed in the classroom. They are also designed to help teachers learn about and understand the whole child. They can help provide information across multiple domains of development and can be sensitive to child growth and development over time. However, to do so, teachers must be collecting valid evidences on an ongoing basis across the entire academic year, not only immediately prior to an assessment checkpoint.

Teachers must also understand that AFA measures ***are not*** benchmark assessments. There ***are not*** correct and incorrect answers to a set of questions or test items, and AFA measures ***are not*** designed to indicate which children are or are not on track to achieve specific summative assessment scores at a specific fixed future assessment date. Rather, AFA measures help teachers understand the developmental status of children wherever the children are developmentally. Each developmental progression can include a wide range of behavioral and observational anchors that extend above and below each age expectation level so as to include opportunities to document child growth and development for all children within the intended age ranges.

Naturally, teachers must understand that AFA measures ***are*** authentic assessments. Authentic assessment resources help teachers observe the progress children are making through a process of gathering evidences of learning that emerge naturally from within daily classroom activities. These evidences are intended to be gathered within regularly occurring instructional activities and routines. Authentic assessment ***is not*** direct assessment. Direct assessments include standardized protocols of assessment activities "done to" a child. This means that for direct assessment measures, children are presented with specific assessment prompts or question formats that are designed to elicit specific correct or incorrect responses from children. While direct assessments can uncover aspects of a child's problem solving processes, direct assessment takes place in an intentionally created artificial testing situation, rather than in the course of daily activities. Direct assessments are appropriate measures for some testing purposes and are widely and correctly used within the broader educational system, particularly with children older than the early childhood years. They can play important roles within a comprehensive assessment system and are appropriately used when objective, summative, data are required concerning how individual children or groups of children are functioning at a particular point in time. Furthermore, direct assessments focus on measurable constructs for which test items can elicit construct-relevant behaviors.

Teachers need to understand that AFA measures ***are*** inherently criterion referenced measures that assess progress and learning relative to a fixed set of standards. They ***are not*** designed specifically to spread out children relative to each other along a continuum of achievement at particular point in time. Rather, they are designed to place children along a continuum of growth and development. The information provided by AFA is most useful for identifying where a given child is functioning relative to their own past developmental trajectory

and relative to standards for children of a given age range. Teachers can use AFA information to understand what behaviors and skills children of a certain age can generally be expected to demonstrate in the classroom.

Criterion referenced assessment tools are not norm referenced tools. AFA generally have not been designed and validated primarily to indicate where a specific child is functioning relative to all other children of similar age. For example, percentile scores are not the focus of the information provided to teachers. Normative information can be made available to teachers as an additional interpretation resource, and can provide general information for teachers who are interested in a broad and comprehensive picture of how a child is growing and developing relative to the developmental progress of other children of similar characteristics, the primary focus needs to remain on specific skills, abilities, and developmental steps.

In contrast to direct assessments, AFA can help teachers follow and examine richly the whole child across a variety of developmental domains as learning and development is unfolding. However, AFA relies on teacher skill and professional judgement as applied to the analysis of a rich portfolio of evidences and documentation of experiences with children across a wide variety of classroom situations and circumstances. Therefore, there is no formal administration protocol for AFA. Rather, as with all authentic assessments, administration is an ongoing process through which teachers observe children in their natural classroom environment, and collect work samples, artifacts, evidences, and anecdotal records that describe and illustrate child learning and developmental progress. After messaging about the purpose of AFA has been developed and delivered, then teacher training and support can begin to focus on observational skills, collecting evidences, analysis of evidences, and making valid placements on the progressions supported by that analysis.

The most important information that results from the use of AFA measures is found at the level of individual placements on the developmental progressions. The current status of a particular child on a particular developmental progression is useful for identifying and describing what the child knows and can do at a particular point time in relation to a specific instructional objective. This information can be used to determine the next targeted developmental milestones for the child and therefore, direct the teacher to the skills and abilities that the child will be developing next. With this knowledge, a teacher can support the developmental progress of the child in a very targeted and individualized way. These placements, or current status levels, can be thought of as raw scores at the item level.

Teachers can evaluate and assess a child's knowledge, skills, and abilities in relation to a particular instructional objective by using the particular developmental progression that is related to that instructional objective. Whenever a teacher has collected sufficient valid evidence regarding a child's current status, the teacher can select the appropriate level along a progression of development and learning. Part of the validity of this information rests in its correspondence to the true ability that a child possesses relative to the particular instructional objective.

Furthermore, at their core, AFA measures are criterion referenced. This means that they are most useful when the information they provide can be interpreted relative to expected levels of development and growth for children of a particular age or grade level. Teachers need a thorough understanding of the ranges of placements that describe the skills, knowledge, and abilities that children of a particular age or class/grade typically demonstrate over a given year of life or from the beginning to the end of a program year. Such ranges of ratings enable teachers to compare data for specific children, and groups of children, and determine if the children's skills, knowledge, and abilities are below, meeting, or exceeding age appropriate expectations.

However, it can be very challenging, stressful, and time consuming for teachers to interpret these ranges of placements to truly differentiate and individualize instruction for all children in their classroom across dozens of instructional objectives and developmental progressions. Teacher training programs often do not adequately prepare teachers for the process of differentiating instruction. Without a well-formed belief system about its value, and specific experiences that include modeling and discourse, teachers often abandon attempts to differentiate in their classrooms (Dack, 2017). However, by understanding the domains of development where individual children need the most support, teachers can not only begin to recognize the value of differentiation and individualization, but can direct their efforts to the areas of greatest need. They can then examine the progression-specific raw score information for those domains of development where overall progress is below expectations. In this way scale scores are a helpful resource to teachers as they use their limited time to triage the needs of children. They can build on the child's areas of relative strength to support growth in the areas needing support (McMillan, 2014).

Valid placements and inferences require teachers to select a sample of evidences that is representative of the child's ability across an entire construct domain. They also require teachers to analyze evidences use that analysis to support make ratings that are truly representative of the child's ability. We can express these concerns through a series of questions about both the rater and the evidences. Would a child be placed at the same level on the developmental progressions if they were in a different classroom or had a different rater? Would a child be placed at the same place on the developmental progressions if they were from a different racial or ethnic group? Would a child be placed at the same level on the developmental progressions if their race or

ethnicity was congruent or incongruent with that of their teacher? Would a child be placed at the same level on the developmental progressions if different evidences were selected?

These questions underscore the need to train teachers to pass inter-rater reliability checks to a sufficient standard of performance. Teacher ratings form the heart of any score information that AFA measures offer. Any child measure dependent on observer ratings provides information about both the raters and the children. This does not mean that the information is merely subjective, and therefore should be excluded from the validity discussion. However, it does mean that evidence of inter-rater reliability is closely connected to, and a necessary component of the validity argument. Once teachers have received and understood appropriate messaging about the nature and purpose of AFA, and have been trained on all the complex concepts and tasks outlined above, then and only then are they ready to meet a standard of agreement with a master rater. This agreement evidence is fundamental for building the argument that the placements on the developmental progressions are not dependent upon who the rater is, and do in fact correspond, to the child's true ability. High quality AFA measures include a process for inter-rater reliability certification as part of the training, and this process can produce some of the most important evidence of implementation fidelity.

These questions also highlight the need to investigate whether a given group of teachers can use an AFA measure to make valid placements for all sub-groups of children. High quality AFA measures are essential to ensure that all children regardless of culture, language, or disabilities are assessed fairly (Qi & Marley, 2009). Furthermore, high quality AFA measures need to be developmentally appropriate, relevant to instruction, and linguistically and culturally responsive (Copple & Bredecamp, 2009). Therefore, evidence that AFA measures yield information that is both equally representative of the true scores of all subgroups of children, and

is equally useful to teachers across all subgroups of children, is an essential component of the construct validity argument (Kim, Lambert, & Burts, 2013).

The relationships between the scores from objective direct assessments and AFA measures can also have a practical use for teachers and those who support them. It can be very helpful for teachers to understand detailed alignments between the information provided by objective direct measures and the indicators from the developmental progressions they are using. The scores children produce when assessed through the artificial situations associated with direct objective measures do not strictly conform to the definition of "authentic" evidence. They are not collected in the regular course of instructional activities. However, teachers have the opportunity to observe the testing situations very closely and can be trained to become sensitive to when the direct testing situation may not be valid for a particular child. Therefore, information from objective direct assessments can constitute valuable evidence to support ratings on the progressions and thereby support child learning, growth, and development. Teachers who have learned to incorporate the formative assessment process into their instruction can use data from direct objective measures as simply another source of evidence to make a child portfolio of evidences more complete. A more complete picture of what a child knows and can do only help the teacher support the child more completely, particularly when the teacher understands exactly how that evidence aligns with child learning standards, instructional objectives, and the steps on the developmental progressions.

Conclusion

Construct-irrelevant variance in assessment scores, no matter what the source, should not result in adverse consequences for any children to whom scores have been assigned (Messick, 1995). Score interpretation should facilitate an accurate understanding of child ability, so that

any decisions that are supported by assessment scores do not result in harmful outcomes for any individual child or subgroups of children. Construct-irrelevant variance in educational assessment scores can result in under-estimation or over-estimation of child ability, which can lead to misclassification, missed opportunities for instructional support and access to helpful resources, or placement in inappropriate educational settings.

AFA measures are not appropriate sources of information to support any high stakes decisions about children or teachers. They are simply not designed or validated to be the single source of information to support placements or classifications of young children, or to inform high stakes evaluation of teachers, schools, or programs. If the information from AFA measures is used for such purposes, there is a clear risk of misinterpretations of scores and harmful consequences for children. Any use of the information from AFA measures needs to adhere strictly to the principles contained in our earlier discussion of the purposes of AFA. Therefore, careful ongoing monitoring of the use of AFA score information is needed to ensure that the information they provide continues to be valid as long as the measure is in use.

Neither the accuracy nor complete utility of information from AFA measures can be realized before a teacher gains a full understanding of, and ability to apply, all of the complex tasks central to the AFA process. Even when teachers have been thoroughly trained and have passed inter-rater reliability, their placements on the progressions are valid only as instructional resources and cannot be treated as if they were scores from objective direct measures. Therefore, interpretations of the information that AFA measures provide are most appropriately utilized when confined to classroom support of the learning process.

For AFA measures, construct-irrelevant variance in the scores can emerge during all phases of the assessment process. If there is not sufficient content coverage within a construct

domain, or if the content of specific developmental progressions does not include a sufficient range of behavioral indicators, teachers may fail to notice or document valid evidences, particularly as manifested by children from historically marginalized groups or subcultures. Therefore, children may not have the opportunity to show their true ability levels which can result in misapplication of differentiated instructional strategies. Between-teacher variance that is not related to variability in the true scores of young children can enter the distribution of scores as teachers determine when they have sufficient evidences, analyze evidences, make placement on the progressions, and adapt instruction based on the placements. Therefore, implementation fidelity evidence to support the use of AFA measures needs to include research evidence rooted in the close monitoring of the extent to which teachers make reasonable placements on the progressions that are supported by valid evidences, correspond to child ability, and are useful to guide effective instructional differentiation for all children.

AFA measures are designed to be instructional resources for teachers and children. Therefore, usefulness to teachers in support of the learning process is the overarching principle that undergirds the validity of the information that AFA measures provide. In this section, we have described the complex set of skills that teachers must master to realize fully the benefits of the FAP in their classrooms. These tasks put real burdens on teachers. Teachers think in terms of the cognitive balancing act between the demands that AFA measures place on their time and energy vis-à-vis the utility of the information they provide. If they appraise that value of AFA information as a resource outweighs its cost, they will more fully engage the process. Value as a resource rests in whether an assessment measure adds to their ability to understand and support instructionally relevant differences between children, create instructionally meaningful subgroups, and plan small group and individual activities.

If the theory behind the AFA process does in fact correspond to the realities of the early childhood classroom, then certain predictions follow. Children can be expected to make greater gains in learning and development in classrooms where teachers implement the FAP with fidelity. In classrooms where teachers are more fully implementing AFA, relative to classrooms where teachers are not as advanced in their use of assessment data as an instructional resource, children can be expected to grow at faster rates. Similarly, teachers in those classrooms should report lower stress levels than their colleagues who view AFA as simply another administrative demand and do not use the information it provides as an essential and well-integrated resource within the instructional process.

References

Akers, L., Del Grosso, P., Atkins-Burnett, S., Monahan, S., Boller, K., Carta, J., & Wasik, B. A., (2015). *Issue Brief—Tailored teaching: The need for stronger evidence about early childhood teachers' use of ongoing assessment to individualize instruction.* Washington, D.C.: Mathematica Policy Research. Retrieved from http://www.mathematicampr.com/~/media/publications/pdfs/earlychildhood/tailored_teaching_child_assessment. pdf.

Baddouh, P., Lambert, R., & Ferrara, A. M. (2016). *An examination of electronic evidences collected as part of a statewide kindergarten entry formative assessment.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Bredekamp, S. (2011). *Effective practices in early childhood education: Building a foundation.* Upper Saddle River, N.J.: Pearson.

Clements, D., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach.* New York: Routledge.

Coburn, C., & Talbert, J. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education, 112*(4), 469-495. doi:10.1086/505056

Cohen-Vogel, L. (2011). Staffing to the test: Are today's school personnel practices evidence based? *Educational Evaluation and Policy Analysis, 33,* 483–505. doi:10.3102/0162373711419845

Copple, C. & Bredecamp, S. (Eds.). (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8.* Washington, D.C.: National Association for the Education of Young Children.

Dack, H. (2017). Structuring teacher candidate learning about differentiated instruction through coursework. *Teaching and Teacher Education, 69*, 62-74. https://doi.org/10.1016/j.tate.2017.09.017

Ferrara, A. M., & Lambert, R. G. (2015). *Findings from the 2014 North Carolina Kindergarten Entry Assessment pilot.* Charlotte: University of North Carolina at Charlotte, Center for Educational Measurement and Evaluation.

Ferrara, A. M., & Lambert, R. G. (2016). *Findings from the 2015 statewide implementation of the North Carolina K–3 formative assessment process: Kindergarten Entry Assessment.* Charlotte: University of North Carolina at Charlotte, Center for Educational Measurement and Evaluation.

Ferrara, A. M., Lin, V., & Lambert, R. G. (2017). *Findings from the 2016-17 North Carolina K-3 formative assessment process implementation case study.* Charlotte: University of North Carolina at Charlotte, Center for Educational Measurement and Evaluation.

Fixsen, D. L., Blase, K. A., Metz, A., & Van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children, 79*(2), 213-230.

Fixsen, D. L., Blase, K. A., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice, 19*(5), 531-540.

Harvey, H. A., & Ohle, K. A. (2018) What's the purpose? Educators' perceptions and use of a state-mandated kindergarten entry assessment. *Education Policy Analysis Archives, 26*(142), 1-33. doi:10.14507/epaa.26.3877

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. doi:10.3102/003465430298487

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin Press.

Heritage, M. (2013). *Formative assessment in practice: A process of inquiry and action.* Boston: Harvard Education Press.

Joint Committee on the Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (AERA/APA/NCME, 2014). *The standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Kim, D-H., Lambert, R. G., & Burts, D. C. (2013). Evidence of the validity of Teaching Strategies GOLD® assessment tool for English language learners and children with disabilities. *Early Education and Development, 24*, 574-595.

Lambert, R. (2018). *Examining the psychometric properties of the North Carolina Kindergarten Entry Assessment* (CEMETR-2018-01). Charlotte: University of North Carolina at Charlotte, Center for Educational Measurement and Evaluation.

Lambert, R. (2019). Shaping a validity argument for the use of authentic formative assessments to support young children. In C. Martin, D. Polly, & R. Lambert (Eds.), *Handbook of research on formative assessment in pre-k through elementary classrooms* (pp. 49-69). Hershey, PA: IGI Global.

Lambert, R. G., Kim, D-H., & Burts, D. C. (2013). Using teacher ratings to track the growth and development of young children using the Teaching Strategies GOLD® assessment system. *Journal of Psychoeducational Assessment, 32*(1), 27-39. doi:0734282913485214

Lesaux, N. K., & Marietta, S. H. (2012). *Making assessment matter: Using test results to differentiate reading instruction.* New York: Guilford Press.

Little, M., Cohen-Vogel, L., Sadler, J., & Merrill, B. (2019). Data-driven decision making in early education: Evidence from North Carolina's pre-k program. *Education Policy Analysis Archives*, *27*(18), 1-23. doi:10.14507/epaa.27.4198

Little, M., Cohen-Vogel, L., Sadler, J., & Merrill, B. (2020). Moving kindergarten entry assessments from policy to practice: Evidence from North Carolina. *Early Education and Development*, 1-21. doi:10.1080/10409289.2020.1724600

McManus, S. (2008). *Attributes of effective formative assessment*. Paper prepared for the Formative Assessment for Student and Teachers (FAST) and State Collaborative on Assessment and Student Standards (SCASS) of the Council of Chief State School Officers (CCSSO). Washington DC: CCSSO; CCSSO FAST SCASS.

McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards based instruction* (6th Ed.). Boston: Pearson.

Merrill, B., Cohen-Vogel, L., Sadler, J., Little, M., & Lee, K. (2017). *A portrait of PreK quality among the American states*. Annual meeting of the Association of Education Finance and Policy. Washington, D.C.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. http://dx.doi.org/10.1037/0003-066X.50.9.741

NAEYC & NAECS/SDE (2003). *Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age eight.* A Joint Position Statement of the National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). Washington, D.C.: Author.

North Carolina K-3 Assessment Think Tank. (2013). *Assessment for learning and development in K-3: A report by the state assessment think tank.* Raleigh, NC: Retrieved from http://www.ncpublicschools.org/docs/newsroom/news/2013-14/20131009/thinktank.pdf

Ohle, K. A., & Harvey, H. A. (2019). Educators' perceptions of school readiness within the context of a kindergarten entry assessment in Alaska. *Early Child Development and Care, 189*(11), 1859-1873. doi:10.1080/03004430.2017.1417855

Qi, C., & Marley, S. (2009). Differential item functioning analysis of the Preschool Language Scale—4 between English-speaking Hispanic and European American children from low-income families. *Topics in Early Childhood Special Education, 29*(3), 171-180. https://doi.org/10.1177/0271121409332674

Roehrig, A. D., Duggar, S., Moats, L., Glover, M., & Mincey, B. (2008). When teachers work to use progress monitoring data to inform literacy instruction: Identifying potential supports and challenges. *Remedial and Special Education, 29*(6), 364-382. doi:10.1177/0741932507314021

Schachter, R. E., Strang, T. M., & Piasta, S. B. (2019). Teachers' experiences with a state-mandated kindergarten readiness assessment. *Early Years, 39*(1), 80-96. doi:10.1080/09575146.2017.1297777

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.

Shepard, L. A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice, 28*(3), 32–37.

William, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53- 84). Mahwah, NJ: Lawrence Erlbaum Associates.

Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 112*(4), 521-548. doi:10.1086/505058

Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. *Education Policy Analysis Archives, 18*(19), 1-38.