Real World Compromises:

Policy and Practice Impacts of Kindergarten Entry Assessment-Related Validity and Reliability Challenges

Debra J. Ackerman

POLICY INFORMATION REPORT

ETS®

Measuring the Power of Learning.®

This Policy Information Report was written by:

Debra J. Ackerman
Educational Testing Service, Princeton, NJ

Policy Information Center
Mail Stop 19-R
Educational Testing Service
Rosedale Road Princeton, NJ 08541-0001
(609) 734-5212
pic@ets.org

Copies can be downloaded from:www.ets.org/research/pic

About ETS

At ETS, we advance quality and equity in education for people worldwide by creating assessments based on rigorous research. ETS serves individuals, educational institutions and government agencies by providing customized solutions for teacher certification, English language learning, and elementary, secondary and postsecondary education, and by conducting education research, analysis and policy studies. Founded as a nonprofit in 1947, ETS develops, administers and scores more than 50 million tests annually — including the *TOEFL*® and *TOEIC*® tests, the *GRE*® tests and *The Praxis Series*® assessments — in more than 180 countries, at over 9,000 locations worldwide.

Policy Information Report and ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Real World Compromises: Policy and Practice Impacts of Kindergarten Entry Assessment-Related Validity and Reliability Challenges

Debra J. Ackerman

Educational Testing Service, Princeton, NJ

Kindergarten entry assessments (KEAs) have increasingly been incorporated into state education policies over the past 5 years, with much of this interest stemming from Race to the Top—Early Learning Challenge (RTT-ELC) awards, Enhanced Assessment Grants, and nationwide efforts to develop common K–12 state learning standards. Drawing on information included in RTT-ELC annual progress reports, published research, and a variety of other KEA-relevant documents, in this report, I share the results of case studies of 7 recently implemented state KEAs. The focus of this inquiry was the assessment- and teacher-related validity and reliability challenges that contributed to adjustments to the content of these measures, the policies regarding when they are to be administered, and the training and technical support provided to teachers who are tasked with the role of KEA assessor and data interpreter. Although these 7 KEAs differ, the results of the study suggest that they experienced common validity and reliability issues and subsequent policy and practice revisions. These findings also suggest the value of iterative research as a means for highlighting these issues and informing the policies and practices that can impact KEA validity and reliability.

Keywords Kindergarten entry assessments; assessment policies; assessment implementation

doi:10.1002/ets2.12201

Over the past 5 years, states have increasingly incorporated kindergarten entry assessments (KEAs) into their education policies, with at least 40 states and the District of Columbia currently developing, piloting, field testing, or implementing these measures (The Center on Standards & Assessment Implementation, 2017). Recent interest in KEAs, which generally are used to inform kindergarten teachers' instruction and state policy decisions, stems from 20 states' federal Race to the Top—Early Learning Challenge (RTT-ELC) awards. Also contributing are Enhanced Assessment Grants awarded to two consortia comprising a total of 17 states (including some RTT-ELC states) as well as to one additional state (Early Learning Challenge Technical Assistance [ELCTA], 2016). In addition, efforts to develop common K–12 state learning standards and to align Grade 3–12 assessments across the United States has brought about a renewed interest in ensuring that kindergarten teachers and education policy makers have accurate information regarding incoming students' academic and developmental skills and needs (Feeney & Freeman, 2014; Goldstein & Flake, 2016).

Despite this collective focus, both the KEAs currently used and their content are not uniform (Weisenfeld, 2017a). Such variations are undoubtedly due to state control of K–12 education (Gottfried, Stecher, Hoover, & Cross, 2011) and differing standards for what kindergartners should know and be able to do (Scott-Little, Kagan, Reid, Sumrall, & Fox, 2014). However, similar to the contextual issues that can shape state pre-K policies and practices (Ackerman, Barnett, Hawkinson, Brown, & McGonigle, 2009), these variations also likely arise from the "real world" compromises that need to be made to bolster an assessment's potential validity, reliability, and utility for specific purposes, populations, and settings (Goldfeld, Sayers, Brinkman, Silburn, & Oberklaid, 2009; Heywood, 2015; Snow & Van Hemel, 2008).

Cross-state research conducted thus far on KEAs has been limited to the measures states mandate teachers to use (e.g., Connors-Tadros, 2014) and some of the implementation issues experienced as they have been piloted and field tested (e.g., Golan, Woodbridge, Davies-Mercier, & Pistorino, 2016). However, because the administration of KEAs is now under way in many states, there is the opportunity to take a retrospective look at the on-the-ground, validity- and reliability-relevant issues that contributed to policy and practice adjustments. By exploring these factors across different KEA contexts, there is the potential to better understand the shaping role they can play and to highlight the value of

Corresponding author: D. J. Ackerman, E-mail: dackerman@ets.org

validity- and reliability-related research for informing assessment policies and practices (Chalhoub-Deville, 2016). Such information could be especially useful as states continue to use KEAs to inform a variety of policy-making purposes (ELCTA, 2015b), but presumably without the financial support of RTT-ELC and other federal funds (Weisenfeld, 2017b).

In this report, I share the results of case studies of seven state KEAs. The focus of this research was some of the assessment- and teacher-associated validity- and reliability-relevant issues that came to light while each KEA was developed, piloted, field tested, and brought to scale, as well as the resulting modifications made to the content of these measures, their administration timelines, and the supports provided to the teachers who serve as KEA assessors, observers, and data users. To set the stage for the study's results, I begin by taking a closer look at the definitions of a KEA and review the concepts of validity and reliability. I then provide an overview of some key psychometric and "real world" classroom implementation factors that can impact an assessment's validity or reliability for a particular purpose and student population. I conclude the report by sharing some implications for early childhood education policy makers and researchers.

What Is a Kindergarten Entry Assessment?

At first glance, the term *kindergarten entry assessment* may appear to be self-defining. However, the term alone does not explain exactly when kindergartners are to be assessed (i.e., prior to kindergarten, the first day of kindergarten, or some other time) or the data to be collected (e.g., children's demographics, home language, or knowledge and skills in a specific academic content area). Most importantly, the term does not indicate the purpose for collecting data (e.g., determine children's eligibility for kindergarten, screen for a potential disability, or inform policy makers' decisions). As a result, it is difficult to understand the theory of action (Bennett, 2011b; Bennett, Kane, & Bridgeman, 2011; Wylie, 2017), or premise, driving the use of KEAs.

Education-focused professional organizations have not coalesced around a single KEA definition, but an examination of their respective descriptions shows some general convergence. For example, the Education Commission of the States (2014) argued that KEAs "evaluate whether or not a student is prepared for the demands of kindergarten." The National Head Start Association (2016) defined KEAs as "assess[ing] what children know and are able to do as they enter kindergarten." The Build Initiative (2016) expanded this definition by defining KEAs as a "process [that] is an organized way to learn what children know and are able to do, including their disposition toward learning, when they enter kindergarten and/or at other times." KEA data also have been touted as an effective way to provide parents, teachers, school and district administrators, and state policy makers with this information (Stedron & Berger, 2010).

The RTT-ELC competition—a major driver in the recent development and revision of KEAs—provides even more details about this type of measure:

> Is administered to children during the first few months of their admission into kindergarten; covers all Essential Domains of School Readiness; is used in conformance with the recommendations of the National Research Council (Snow & Van Hemel, 2008) reports on early childhood; and is valid and reliable for its intended purposes and for the target populations and aligned to the Early Learning and Development Standards. Results of the assessment should be used to inform efforts to close the school readiness gap at kindergarten entry and to inform instruction in the early elementary school grades. This assessment should not be used to prevent children's entry into kindergarten. (U.S. Department of Education, 2011)

Similarly, the Enhanced Assessment Grant KEA competition defined these measures as

> provid[ing], at kindergarten entry, valid and reliable information on each child's learning and development across the essential domains of school readiness … [and] would be used to support educators in providing effective learning opportunities to every child, and help close achievement gaps. (U.S. Department of Education, 2013)

Although the RTT-ELC and Enhanced Assessment Grant definitions provide similar guidance about the purpose of a KEA and when it should be administered, neither definition specifies the assessment approach to be used (i.e., direct measure with select responses vs. a measure that relies on an observational rubric and requires teachers to collect evidence of a student's knowledge or skill) or the exact items to be included. As a result, states receiving these funds have had significant

leeway to select or develop a KEA that best meets their needs. Indeed, the state-mandated and -recommended KEAs used across the United States include commercially available measures, newly developed assessments, and state-developed instruments. In addition, these measures represent both direct and observational approaches (Weisenfeld, 2017a).

At the same time, both the RTT-ELC and Enhanced Assessment Grant definitions stated that KEAs must be "valid and reliable," a key factor that policy makers and other education stakeholders need to bear in mind when developing or selecting any assessment. *Validity* refers to the extent to which the interpretation of a measure's scores provide sufficient evidence to inform a specific purpose and population of students (Bonner, 2013; Kane, 2013). The focus on purpose and population is important, as the inferences made from these scores can be more or less valid depending on the purpose for which the test is used and the population for which inferences are made. Thus, although validity refers to the degree to which an assessment measures what it is supposed to measure, a key question to ask is how to design a measure so that it is maximally useful for its primary purpose, such as informing a kindergarten teacher's practice or guiding state policy makers' decisions. The term, *reliability* refers to the extent to which a measure provides consistent results over different assessors, observers, testing occasions (assuming there is a minimal time lapse between testing occasions), and test forms (Snow & Van Hemel, 2008). Researchers also have been concerned with the unintended consequential validity of assessments, particularly if test results are being used for important decisions beyond their original purpose (Wylie, 2017). For example, both the RTT-ELC and Enhanced Assessment Grant competitions explicitly stated that a KEA should *not* be used to deny an otherwise age-eligible child's entry into kindergarten, which was a major criticism of the school readiness tests that were prevalent in the 1980s (Bowman, Donovan, & Burns, 2001; Shepard, 1997).

Concerns about inconsistent assessment practices across different education settings can likely be mitigated via policy (e.g., requiring all kindergarten teachers to use a specific KEA within a standardized time period). However, additional validity- and reliability-relevant issues can arise from an assessment itself as well as from the individuals who administer and score it and use its data as evidence. As I discuss next, different types of research studies have the potential to highlight needed modifications to a measure's items, how it is administered and scored, and the policies governing these processes.

Assessment- and Assessor-Related Validity and Reliability Challenges

When an assessment is used in classrooms serving young children, the average parent or education stakeholder might presume that the measure is "good to go" for accomplishing any of the purposes for which the data are being used (Regenstein, Connors, Romero-Jurado, & Weiner, 2017). Yet, what may be less obvious is the inherent difficulty in designing a measure that is valid and reliable for its intended purposes and populations (Lambert, 2003). In fact, researchers have noted that assessment development can involve a significant financial investment, effort, fault tolerance, and timeline (Bennett, 2011a; Pellegrino, Chudowsky, & Glaser, 2001). When one adds the degree to which teachers are — or are not — prepared to administer and use the results of classroom assessments (Campbell, 2013), one can begin to understand that merely enacting a policy mandating the use of a specific measure at a specific point in time does not necessarily mean the goals of conducting the assessment (e.g., inform teachers' instruction) will be realized.

To bolster a measure's validity and reliability for a particular purpose and population, assessment developers and researchers will ideally conduct an array of studies. These studies, which can include a focus on the psychometric properties of an assessment and the different individuals who interact with it (i.e., administrators, test takers, score users), are admittedly complex and thus deserving of their own report. They also will vary based on the measurement approach (e.g., direct assessment or observational rubric), the different individuals and systems involved in the entire testing and data use process, how often an assessment is administered, and the purposes for which the measure is being used (Mislevy, Wilson, Ercikan, & Chudowsky, 2003; Snow & Van Hemel, 2008). Furthermore, recent scholarship has argued that if a theory of action undergirds the use a test (e.g., KEA data will inform teachers' instruction and contribute to policy makers' efforts to close achievement gaps), such research should not only investigate whether that utility hypothesis is being realized but also inform relevant assessment policies (Bennett, 2011b; Bennett et al., 2011; Chalhoub-Deville, 2016).

In light of this larger context, I highlight here seven key validity and reliability issues that can be discerned through various types of research and have particular relevancy to KEAs when used as early childhood formative measures (Riley-Ayers, 2014). As Figure 1 outlines, four of the issues can emanate from an assessment. The remaining three issues are related to assessor or observer capacity.

Assessment-Related Issues
- Alignment with purpose, population, curriculum, and data use timeline
- Individual items focus primarily on a single construct
- Adequate allowances for English learners or children with diagnosed special needs
- Sufficient sensitivity to measure variations in development

Teacher-Related Issues
- Administration/observation capacity
- Timeline pressures
- Access and data use capacity

Figure 1 Potential kindergarten entry assessment validity and reliability issues.

Potential Assessment-Related Issues

Alignment

When developing or selecting an assessment or observational rubric, one initial topic for assessment stakeholders to explore is the extent to which its items appear to be appropriate for generating sufficient evidence for a specific purpose and population. This is a key issue for KEAs, as these measures essentially make claims about the skills and knowledge that are important for young children to develop by the time they enroll in kindergarten (North Carolina K–3 Assessment Think Tank, 2013; North Carolina Office of Early Learning, 2015c). As an example, if a state's KEA will be used, in part, to provide evidence of a kindergarten student's language and literacy skills, those tasked with selecting a measure would want to investigate whether there are items that appear to address these content areas in a way that is developmentally appropriate for children who are roughly 4–7 years old.

To accomplish this appropriate-evidence task, those tasked with selecting or developing an assessment often conduct what is known as a *crosswalk* or alignment study. One common crosswalk focus is the match between a state's learning standards for a specific age or grade and a measure's content domains and items (Cox, Rodriguez, & Edwards, n.d.; Irvin et al., 2012; Roach, McGrath, Wixson, & Talapatra, 2010). If the results of an assessment are to inform teachers' practice, another common alignment focus is whether the knowledge or skills that are focused on are targeted by the curriculum used (National Association for the Education of Young Children & National Association of Early Childhood Specialists in State Departments of Education, 2003; Wakabayashi & Beal, 2015). An additional alignment topic is when teachers and other school stakeholders will have access to score data (Jiban, 2013; Kallemeyn & DeStefano, 2009). In either of these two latter cases, assessment results will have limited utility if they focus on skills that teachers do not teach or an individual student's score data are received long after he or she has moved on to the next grade.

Item Construct Consistency

Although an assessment or observation may appear to be aligned with a particular purpose and population, a second potential issue is the extent to which its individual items consistently measure the constructs that they are intended to measure (Bonner, 2013; Cook & Beckman, 2006; Howard et al., 2017). Multiple factors can contribute to problems in this area. For example, a single item may cover an array of discrete skills but require teachers to assign a single score, resulting in a lack of specificity. Or the item may be worded so vaguely that it can be interpreted in different ways or is dependent on specific situations. If the measure uses an observational approach and scoring rubric, a closely related issue is whether the authors provide sufficient guidance for what constitutes valid and sufficient evidence upon which to base a judgment (Dever & Barta, 2001; Goldstein & McCoach, 2011; Harvey, Ohle, & Leshan, 2017; Snow & Van Hemel, 2008). To help address these specific issues, researchers can conduct *think-alouds* or cognitive interviews to determine how individuals are interpreting text and thus why they are responding to item prompts, or scoring items, in a particular way (Bonner, 2013; Jonsson & Svingby, 2007; Snow & Van Hemel, 2008).

Item construct issues also can arise if children's responses are influenced by, or dependent on, other potentially irrelevant factors, including their receptive and expressive vocabulary levels, or even the time of day or their physical health (Lane, 2013; Najarian, Snow, Lennon, Kinsey, & Mulligan, 2010; North Carolina Office of Early Learning, 2015c). Researchers may therefore examine a measure's relationship with the scores from other similar measures, as such correlations tend to support the belief that the measure of interest is valid for that particular purpose and population (Goldstein, McCoach, & Yu, 2017; Howard et al., 2017; Lai, Alonzo, & Tindal, 2013). They also can examine the consistency of

scores for individual children, or what is known as *test–retest reliability*. That is, if we gave the same direct assessment to a child within a short period of time (e.g., 2 weeks), unless there is reason to expect a change in the child's ability, we would expect his or her scores to be similar (Cook & Beckman, 2006; Muller, Kerns, & Konkin, 2012). Researchers also may wish to ensure that items are measuring the same construct across groups of children by assessing a representative sample of students from different demographic, linguistic, and cultural backgrounds (Bennett, 2011a; Quirk, Rebelez, & Furlong, 2014).

Adequate Allowances

Examining a measure's score validity for all students can be especially critical for English language learner students, as there is the potential to mistakenly interpret low scores on measures in which the students must respond to prompts in English as evidence of inadequate content knowledge (Alvarez, Ananda, Walqui, Sato, & Rabinowitz, 2014; Robinson, 2010). At the same time, assessment developers need to consider the allowances that may be used when students have a home language other than English or diagnosed special needs. As an example, some young English language learners may be better able to demonstrate their content skills (e.g., quantifying or identifying letters, shapes, or other objects) when using their home language exclusively or a mixture of that language and English (Lopez, Turkan, & Guzman-Orth, 2017). However, to support the measure's validity and reliability, the individual administering the test or collecting evidence via an observational rubric presumably would need to be fluent in the child's home language, a point to which I return later.

Although teachers may be tempted to create allowances on an impromptu basis (Gokiert, Noble, & Littlejohns, 2013), such modifications can affect the validity and reliability of assessment data for a specific purpose. For example, informally translated assessment prompts may not be accurate or the substituted words within an item may not have the same level of difficulty as was found in the original assessment (Atkins-Burnett, Bandel, & Aikens, 2012; Golan et al., 2013). Accordingly, the Division for Early Childhood of the Council for Exceptional Children (2010) has advised education stakeholders to use extra caution in interpreting the results of assessments that have been informally translated into a child's language rather than undergoing a more formal research-based translation.

Sufficient Sensitivity

A fourth potential issue to investigate is whether an assessment or observational rubric is sensitive enough to measure variations in development and, in turn, inform the purpose for conducting the assessment or observation. This focus can be due to the need to measure changes over time (e.g., examine the effects of a curriculum or intervention; Garon, Smith, & Bryson, 2014; Karelitz, Parrish, Yamada, & Wilson, 2010) or, in the case of a KEA, to document the wide range of children's skills at a particular point in time (Quirk, Nylund-Gibson, & Furlong, 2013; Snow & Van Hemel, 2008). In either case, if a measure only reflects beginner or advanced skills, it likely will not be appropriate for assessing all of the students in a teacher's classroom (Johnson & Buchanan, 2011; Jonas & Kassner, 2014).

To explore this issue, researchers can investigate whether an assessment suffers from what are known as floor and ceiling effects. Put another way, tasks need to be sufficiently easy for measuring children's skills at the earliest stage of development (the floor) but also adequate for when they are older and might consistently hit the ceiling of the range of possible scores (Dever & Barta, 2001; Rock & Pollack, 2002). This issue can be particularly tricky when assessing entering kindergartners, as students will typically exhibit a wide range of skills on Day 1 but also pick up certain skills like letter naming or the sounds of commonly used letters quite quickly (Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009; Torgeson, 1998). Psychometric analyses also can assess whether assessments are more sensitive to low to average levels versus higher levels of skills and knowledge (Cook & Beckman, 2006; Messick, 1987).

Potential Teacher-as-Assessor/Observer Issues

Capacity to Administer an Assessment or Collect Observational Evidence

Assessment developers also need to be mindful of the needs of the staff who interact with a measure at any stage of the testing process (Snow & Van Hemel, 2008). One key validity and reliability issue to explore is teachers' capacity to administer an assessment as intended by its developers (Parkes, 2013). This issue is especially salient when discussing KEAs,

as many of these measures rely on observational rubrics as opposed to direct assessments and thus may be unfamiliar to the kindergarten teachers who are tasked with administering them (Butts, 2013; Schultz, 2014; Weisenfeld, 2017a). Moreover, the process of making reliable judgments about students' skills based on a rubric or rating scale can be challenging for many teachers (Cabell, Justice, Zucker, & Kilday, 2009; Furnari, Whittaker, Kinzie, & DeCoster, 2016; Kilday, Kinzie, Mashburn, & Whittaker, 2012; Mashburn & Henry, 2005; Waterman, McDermott, Fantuzzo, & Gadsden, 2012). Rubric or rating scale reliability can be affected if the scores are perceived to contribute to some type of consequential decision for the student, teacher, or program, as well (Harvey, Fischer, Weieneth, Hurwitz, & Sayer, 2013; Waterman et al., 2012). While such bias is rarely intentional, it may result in over- or underestimating children's proficiency in any area.

Not surprisingly, score reliability may be negatively impacted by teachers' lack of proficiency with a measure (Jonas & Kassner, 2014; Miller-Bains, Russo, Williford, DeCoster, & Cottone, 2017). It therefore makes intuitive sense that teachers' capacity may be dependent on adequate training (Grisham-Brown, Hallam, & Pretti-Frontczak, 2008; Kamler, Moiduddin, & Malone, 2014; Snow & Van Hemel, 2008). Accordingly, the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (2014), as well as the position statement issued by National Association for the Education of Young Children and National Association of Early Childhood Specialists in State Departments of Education (2003), stress that assessment administrators should be adequately trained and demonstrate proficiency in administering and scoring any measure. For observational measures specifically, observers also should demonstrate that their training has been sufficient for using the measure and its scoring rubric in a way that is consistent with its developer's intention.

Teacher capacity also includes the ability to reliably assess children who are considered to have special needs or are English language learners. The Division for Early Childhood of the Council for Exceptional Children (2007) has urged school staff to have regular access to training on administering, scoring, and interpreting the results of assessments. Similarly, the National Association for the Education of Young Children (2009) advised that English language learners be assessed by well-trained staff who also are bilingual and bicultural. Being bilingual can be important, as some students may be better able to demonstrate their content skills when tested in or using their home language rather than English (Lopez et al., 2017). Furthermore—and as mentioned earlier—when relying on score data from assessments in English to make some type of decision about individual English language learner students, test administrators and observers need to distinguish between inadequate content knowledge and a student's lack of English language proficiency (Alvarez et al., 2014; Robinson, 2010).

No matter the assessment or population being assessed, teachers' administration capacity should not be assumed either prior or subsequent to training. Instead, both intra- and interrater consistency should be confirmed at the conclusion of training and tested repeatedly over time (McClellan, Atkinson, & Danielson, 2012). Establishing appropriate reliability levels promotes consistency within and across classrooms and provides confidence that children's assessment scores reflect their performance on a particular occasion rather than the capacity of the assessor (Wright, 2010). In the case of KEAs specifically, rater reliability is particularly important if the data are also used to inform district- or state-level decisions.

Timeline Pressures

A second potential teacher challenge is the amount of time it can take to administer a KEA or other assessment. Time challenges can stem from a variety of sources, including the number of items to be completed and the type of response needed (e.g., selected response, constructed response, or hands-on task; Davy et al., 2015; Jones & Vickers, 2011). The testing platform (e.g., paper and pencil vs. computer based) also can play a role, particularly if computers, laptops, tablets, testing software, or Internet systems are not working as expected (Martineau, Domaleski, Egan, Patelis, & Dadey, 2015).

Regardless of the source, this issue is salient for KEAs, as teachers generally need to complete these assessments within 30–60 days after the school year begins (ELCTA, 2016). Having "enough time" can be especially problematic if teachers are expected to administer a measure to, or observe, all children in the classroom within a specific time frame (Banerjee & Luckner, 2013; Basford & Bath, 2014; Bradbury, 2014; Susman-Stillman, Bailey, & Webb, 2014; Zellman & Karoly, 2012; Zweig, Irwin, Kook, & Cox, 2015). As an example, if the administration of a KEA requires an average of 1 hour per student, a kindergarten teacher with 30 full-day students can easily use up an entire week of available instructional time to complete the testing or observation process. In fact, in 2014, Florida policy makers dropped a language- and literacy-focused assessment from its suite of KEA measures because of teacher complaints regarding the amount of instructional time lost to administering the measure (Jester, 2014; Strauss, 2014).

The amount of time needed to administer an assessment in and of itself typically does not pose a validity or reliability threat. However, perceptions of insufficient time can contribute to such threats if they cause teachers to deviate from the accepted administration protocol (Kim, 2016). For example, an evaluation of Nevada's 2014 pilot of its observational KEA revealed that teachers were arbitrarily selecting which items to focus on or skip as a means for cutting back on the total amount of time needed to observe students (Loesch-Griffin, Christiansen, Everts, Englund, & Ferrara, 2014). Time pressures also may result in teachers not choosing to administer an assessment (MacDonald, 2007). These prior studies suggested that it can be useful to conduct surveys or focus groups with teachers to ascertain to what extent they are experiencing challenges in assessing their students within the prescribed time period.

Data Access and Use Capacity

A third potential teacher-as-KEA-assessor/observer challenge is teachers' capacity to use the results of assessments to guide their pedagogical decisions (Akers et al., 2014; Ball & Gettinger, 2009; Greenberg & Walsh, 2012; Strand, Cerna, & Skucy, 2007). As mentioned, one potential issue is gaining access to score reports in a timely manner (Jiban, 2013; Kallemeyn & DeStefano, 2009). Teachers also need adequate capacity to access the system housing the data (Roehrig, Duggar, Moats, Glover, & Mincey, 2008).

Even if score reports or anecdotal data are available immediately, teachers do not necessarily have the skills and knowledge needed to make the leap to using these data as intended (Isaacs et al., 2015; Pyle & DeLuca, 2017). In short, "programs may collect data, then not know what to do with the information" (Yazejian & Bryant, 2013, p. 68). This is a particularly salient issue when using formative assessments such as KEAs, as one premise undergirding the use of these measures is that teachers will adjust their instruction based on what they know about students' current skills and knowledge. In this case, even if teachers produced reliable scores, assessment validity can be challenged due to their inability to use the evidence to inform the original purpose for collecting these data.

In summary, the mere selection or use of a KEA (or any other assessment) may not result in valid and reliable evidence to inform any purpose or population. This can particularly be the case if the measure used suffers from item- or scoring-associated issues or teachers do not have appropriate administration, implementation, or score usage capacity. While not every measure will suffer from these challenges, this research base suggests that simply establishing a KEA policy may not be sufficient. Instead, it also could be useful to engage in a continuous plan, use, review, and revise research model (Riley-Ayers, Frede, Barnett, & Brenneman, 2011) to generate data on which inputs are — and are not — working as needed to meet the policy's goals.

The Current Study

KEAs have been increasingly incorporated into most states' educational policies over the past 5 years, but the actual instruments used, the content of these measures, and their administration policies vary widely between states (ELCTA, 2016; Weisenfeld, 2017a). An array of state-controlled factors, including differing K–12 policies (Gottfried et al., 2011) and kindergarten learning standards (Scott-Little et al., 2014), likely contribute to these variations. However, assessments also can be shaped by the "real world" validity and reliability issues that come to light as these measures evolve from initial design or selection to full implementation in a particular setting (Goldfeld et al., 2009; Heywood, 2015).

Research thus far on KEA validity and reliability issues and subsequent policy and practice revisions is limited (e.g., Golan et al., 2016) and likely reflects the fact that federal support to develop these measures only began in 2012. Therefore the current study aimed to determine some of the assessment- and teacher-associated validity and reliability issues that have arisen as these KEAs have been selected/developed, piloted, field tested, and scaled up over the past 5 years and, in turn, contributed to revisions to their (a) content, (b) administration timeline policies, and (c) teacher training on or technical assistance for administering these measures as intended by their developers, accessing KEA data, and using these data to inform instruction.

Method

To address these three focus areas, I used a case study approach. This approach is particularly well suited to investigating some of the "real world" issues that contributed to changes in state KEA policies and practices, because although each

KEA is a bounded unit, they also share similar components (e.g., the measure itself, teacher professional development) and goals (serving as a formative assessment that will inform teachers' practice). Therefore, by treating each KEA as a case, I was able to analyze data both within and across each of them and consider the implications of the study's results as a whole (Merriam & Tisdell, 2016).

Sample

The study's cases were comprised of the seven KEAs used in 2017 in the states of Delaware, Illinois, Maryland, North Carolina, Ohio, Oregon, Pennsylvania, and Washington (see Table 1). I purposefully selected these KEAs for four reasons. First, they represent the different types of KEAs used across the United States in terms of being a customized version of a commercially available measure, newly developed, or a revision of an existing state-developed measure. For example, Delaware and Washington each use customized versions of Teaching Strategies GOLD (Heroman, Burts, Berke, & Bickart, 2010), a commercially available observational assessment that is used in these two states' and other states' publicly funded pre-K programs (Ackerman & Coley, 2012). GOLD also is used as a KEA in seven additional states (Weisenfeld, 2016) but had not previously been used in this capacity in Delaware or Washington. Maryland and Ohio are collaborating on the development of the new Kindergarten Readiness Assessment, which includes both observational and direct items. Illinois, North Carolina, and Pennsylvania elected to enhance existing state-developed kindergarten assessments (U.S. Department of Education & U.S. Department of Health and Human Services, 2014), but with Illinois's measure originally used in California. Although North Carolina also enhanced its state-developed KEA, the measure's online data storage system is a customized version of what is available to GOLD users (Public Schools of North Carolina, 2015). Finally, Oregon's KEA consists of literacy and mathematics items from a commercially available direct assessment developed by researchers at its state university as well as items from the observation-based Child Behavior Rating Scale (Bronson, Goodson, Layzer, & Love, 1990; Bronson, Tivnan, & Seppanen, 1995).

Second — and as also can be seen in Table 1 — these KEAs vary in terms of the domains assessed. All seven KEAs focus on literacy and cognition or mathematics, and six of the seven also focus on children's physical and social–emotional development. However, the total number of domains varies across these measures. Additionally, these measures vary in terms of the amount of time teachers have to assess or observe their students.

Third, I selected these KEAs because of their respective states' receipt of RTT-ELC or Enhanced Assessment Grant awards, as I assumed these funds served as a proxy for states having the motivation and financial means not only to select or develop, pilot, field test, and scale up these measures but also to provide teacher professional development and engage in measure- and assessor-related research aimed at mitigating potential validity and reliability challenges. For example, Illinois's RTT-ELC application outlined the state's plan to select and adapt a high-quality KEA, undertake a phased implementation plan, and provide teachers with professional development on child observation, administering the measure in a reliable manner and using the score data to improve and align their instruction. The application also outlined the state's plans to conduct a validation study (Office of Governor Pat Quinn, 2011).

Finally, I selected these KEAs because of their respective development timelines, which needed to be advanced enough for there to be data available on the results of their measure-related research. The states of Illinois, Pennsylvania, and Washington were selected because they received RTT-ELC awards and phased in implementation of their respective measures by 2016. Similarly, Delaware, Maryland, North Carolina, Ohio, and Oregon received funds from both RTT-ELC and the federal Enhanced Assessment Grant competition to support their KEA work and, by 2016, were fully implementing their respective measures (ELCTA, 2016).

Data Collection and Analysis

To collect data for the case studies and address the study's three focus areas, in the summer of 2017, I gathered documents related to the development, piloting, field testing, and scale up of each of the seven sample KEAs. Documents can serve as useful sources of qualitative data (Yin, 2014), particularly when analyzing legislation based education policies (Gibton, 2016) or tracking changes over time (Bowen, 2009). The documents I used included RTT-ELC annual progress reports, technical reports, research briefs, newspaper and peer-reviewed journal articles, KEA administration manuals, state administrative memoranda, and presentations (see the appendix). These documents were identified and retrieved

Table 1 Sample Kindergarten Entry Assessments

Name/type of kindergarten entry assessment	Domains	Fall timeline
Customized version of Teaching Strategies GOLD		
Delaware Early Learner Survey[a]	Language and Literacy Development Cognition and General Knowledge Approaches Toward Learning Physical Well-Being and Motor Development Social and Emotional Development *For English language learners*: English Language Acquisition	first 30 days of school[b]
Washington Kindergarten Inventory of Developing Skills (WaKIDS)[c]	Language, Literacy Cognitive, Mathematics Physical Social–Emotional	by October 31[d]
New measure		
Maryland[e]/Ohio[f] Kindergarten Readiness Assessment	Language and Literacy Math Physical Well-Being and Motor Development Social Foundations	first day of school to early October (Maryland)[g] November 1[h] (Ohio)
Enhancements of existing state-developed measures		
Illinois Kindergarten Individual Development Survey (KIDS)[i]	Language and Literacy Development Cognition: Math; Cognition: Science History–Social Science Visual and Performing Arts Approaches to Learning: Self-Regulation Physical Development; Health Social and Emotional Development *For English language learners*: Language and Literacy Development in Spanish English-Language Development	within 40 days of attendance[j]
North Carolina Kindergarten Entry Assessment[k]	Language Development and Communication Cognitive Development Approaches to Learning Health and Physical Development Emotion: Social Development	within first 60 days[l]
Pennsylvania Kindergarten Entry Inventory[m]	Language and Literacy Development Mathematics Approaches to Learning Health, Wellness, and Physical Development Social and Emotional Development	within first 45 calendar days[n]
Items from several existing measures		
Oregon Kindergarten Assessment (easyCBM Literacy and Math subtests; Child Behavior Rating Scale)[o]	Early Literacy Early Math Approaches to Learning	within first 6 weeks of kindergarten[p]

[a]https://www.doe.k12.de.us/Page/3029. [b]Delaware Department of Education (DDOE, 2016a). [c]http://www.k12.wa.us/WaKIDS/. In addition to the customized version of the GOLD assessment, WaKIDS also comprises a Family Connection teacher–family meeting and district-level collaboration with students' early care and education providers (Washington Office of Superintendent of Public Instruction, 2017). [d]Dorn (2016). [e]https://pd.kready.org/105956/. [f]http://education.ohio.gov/Topics/Early-Learning/Kindergarten/Ohios-Kindergarten-Readiness-Assessment. [g]Maryland State Department of Education (2017b). [h]Ohio Department of Education (2017); Ohio Department of Education Office of Curriculum and Assessment (2017). [j]Illinois State Board of Education (n.d.). [k]http://nck-3fap.ncdpi.wikispaces.net/home. [l]North Carolina Statute 115C-83.5 (http://nck-3fap.ncdpi.wikispaces.net/home). [m]http://www.education.pa.gov/K-12/Assessment%20and%20Accountability/Pages/Kindergarten-Entry-Inventory.aspx#tab-1. [n]Pennsylvania Office of Child Development and Early Learning (2015). [o]http://www.oregon.gov/ode/educator-resources/assessment/Pages/Kindergarten-Assessment.aspx. [p]Office of Assessment and Accountability, Oregon Department of Education (2016b).

through Internet searches and from presentations by the researchers involved in developing or evaluating these measures and their respective rollouts.

To analyze the data in these documents, I followed the approach advocated by Bowen (2009). I first read through the KEA–/state-specific documents in chronological order (oldest to most recent) to note any assessment- and assessor-associated issues and resulting modifications made over time. I then read these documents more closely so that I could code the issues and revisions using the assessment (e.g., alignment, floor/ceiling, content feedback issues) and assessor

(e.g., training, time pressures, capacity to access and use results) categories highlighted in my literature review. Because most of these issues and resolutions were mentioned across multiple documents, I also was able to triangulate this information.

All of the coded information was recorded in an Excel database that had a row for each KEA and a column for each issue code. The individual cells contained information about the issue, its eventual resolution, and the documents that contained this information. By setting up the database in this way, I could determine the array of issues and adjustments within each KEA and the extent to which these issues and adjustments were experienced across the entire sample. These cross-sample results are presented next.

Results

Issues Resulting in the Revision of Kindergarten Entry Assessment Content

The first focus of my study was the assessment- and teacher-related issues that prompted revisions to the content of states' KEAs. Analysis of the study's document-based data demonstrated all seven KEAs experienced such revisions. Furthermore—and as is displayed in Table 2—these revisions were prompted by efforts to align learning standards with KEA content, floor and ceiling issues, feedback on content, and "time crunch" pressures.

Alignment Issues

The development of Illinois's Kindergarten Individual Development Survey (KIDS; Illinois State Board of Education, 2015), an observational measure that uses a developmental progression-related scoring rubric, provides an example of the impact of a state's kindergarten learning standards on the expansion of a KEA. KIDS is modeled on the Desired Results Developmental Profile: School Readiness (DRDP-SR) measure, which was originally developed by the California Department of Education Child Development Division (2012). In 2012, the DRDP-SR included the domains of Self and Social Development, Self-Regulation, Language and Literacy Development, Mathematical Development, and English Language Development and thus was aligned with the 2010 Kindergarten Common Core State Standards and WIDA (2012) standards for linguistically diverse students (WestEd Center for Child and Family Studies Evaluation Team, 2013). However, these five domains did not adequately represent all of Illinois's kindergarten learning standards (Illinois State Board of Education Division of Early Childhood Education, n.d.). As a result, in collaboration with the Illinois State Board of Education, both KIDS and a new kindergarten version of the DRDP were expanded to include 11 domains, 2 of which are conditional for English language learners (see Table 1). Items also were added to some of the existing domains (California Department of Education Child Development Division, 2015; Kriener-Althen, Burmester, & Mangione, 2015).

Alignment with early learning standards also played a role in the revisions made to North Carolina's Kindergarten Entry Assessment, which uses an observational approach to document children's skills and knowledge across 10 "construct progressions" (subdomains). This state-developed KEA originally focused on reading and mathematics skills. It therefore was expanded to be aligned with the learning standards in the domains of Language Development and Communication, Cognitive Development, Approaches to Learning, Health and Physical Development, and Emotion-Social Development (Office of the Governor, State of North Carolina, 2013, 2014).

Similarly, when first piloted in 2011, Pennsylvania's KEA, which also is an observational measure, contained the domains of Social Emotional, Language and Literacy, Mathematical Thinking, and Approaches to Learning. In fact, the

Table 2 Issues Resulting in the Revision of Kindergarten Entry Assessment Content

Kindergarten entry assessment	Alignment	Floor/ceiling	Content feedback	Time crunch
Delaware Early Learner Survey				X
Illinois Kindergarten Individual Development Survey (KIDS)	X	X		
Maryland/Ohio Kindergarten Readiness Assessment			X	X
North Carolina Kindergarten Entry Assessment	X		X	
Oregon Kindergarten Assessment	X	X	X	
Pennsylvania Kindergarten Entry Inventory	X		X	
Washington Kindergarten Inventory of Developing Skills (WaKIDS)				X

measure's name at that time—SELMA—reflected the acronym of these four domains (Pennsylvania Office of Child Development and Early Learning [Pennsylvania OCDEL], 2013). However, to align with the state's kindergarten learning standards, four additional health, wellness, and physical development items were incorporated during the 2012 pilot. In addition (and presumably because the SELMA acronym was no longer appropriate), the name of the measure was changed to the Kindergarten Entry Inventory (Pennsylvania OCDEL, 2014a).

Floor and Ceiling Issues

The development of Oregon's Kindergarten Assessment provides an example of the impact of psychometric research on a measure's items. The state's KEA comprises items from the easyCBM (Anderson et al., 2014), a direct assessment, and the Child Behavior Rating Scale (Bronson et al., 1990; Bronson et al., 1995), which relies on teacher observations of students to assess their approaches to learning. The 2012 pilot version of the Kindergarten Assessment also included the easyCBM phoneme segmentation fluency items. However, these items were dropped after the pilot because of analyses demonstrating that they were too difficult for the majority of entering kindergartners (Furrer & Green, 2013; Irvin, Tindal, & Slater, 2017; Tindal, Irvin, Nese, & Slater, 2015).

Similar research played a role in the revision of Illinois's observational KIDS measure. Researchers examining children's scores during the 2013–2014 field testing of the measure discovered that 5%–10% of kindergartners were being rated at the highest level for most of the learning-related items. In response, the California-based assessment developers added a sixth, more advanced developmental level to the scoring rubric (Kriener-Althen et al., 2015; Office of the Governor, State of California, 2016).

Feedback on Content

More detailed revisions to North Carolina's KEA took place after the measure was used at the beginning of the 2015–2016 school year, and when teachers were required to use just 3 of the 10 construct progressions due to administration time issues (and highlighted in the next section). Following this administration, teacher feedback indicated the need for further clarity in some of the skill text and performance descriptions that were used to determine children's developmental levels within these three progressions. In response, in 2016, the state's Office of Early Learning revised this text. For example, in the Book Orientation construct progression, one item is "Children understand that books have pages that contain pictures and/or words." One included skill that demonstrated this understanding was the holding and opening of a book as well as turning the pages. To provide more clarity, the wording specific to "turns pages" was revised to "turns pages front to back" (North Carolina Office of Early Learning, 2017a; Pruette, 2017).

Similarly, the use of think-alouds during the development of Maryland and Ohio's Kindergarten Readiness Assessment impacted the measure's initial content. This KEA consists of observational items as well as selected-response and performance-based direct items. Owing to feedback from January 2013 cognitive interviews, as well as the results of a post-April 2013 pilot test administrator questionnaire, the measure's assessment developer revised item types, content, wording, graphics, and administration procedures (Office of the Governor, State of Maryland, 2014; Office of the Governor, State of Ohio, 2014; WestEd, 2014).

Teacher feedback, as well as standards alignment work, further impacted Oregon's KEA. Originally, the English letter naming and letter sound items were timed, with children's scores based on the number of correct answers provided in 1 minute. In 2016, these items were replaced by untimed upper and lowercase letter name and letter/sound recognition items. In addition, the original timed Spanish Letter Names subtest was revised to an untimed Spanish Letter Name and Sounds subtest.[1] The Spanish version of the mathematics portion of the assessment was made available as well (Early Learning Division, Oregon Department of Education, 2016, 2017; Office of Assessment and Accountability, Oregon Department of Education, 2016a, 2016b; Office of Teaching, Learning, and Assessment, Oregon Department of Education, 2017; Oregon Department of Education, 2015b, 2017b).

Teacher feedback also contributed to further revisions of Pennsylvania's pilot SELMA measure, which was the precursor to the state's current Kindergarten Entry Inventory. Initially, this observational measure's rubric had four developmental categories: not yet evident, emerging, evident, and exceeds. However, focus group feedback from 2011 pilot teachers indicated that these categories did not provide an option when there was no opportunity to observe a specific skill during the administration period or when a specific skill would be inappropriate for a particular child (e.g., due to a disability).

As a result, in 2012, the state added a fifth "unable to determine" level. Teachers also are required to note the reason why they were unable to observe a specific indicator (Pennsylvania OCDEL, 2014a, 2014b, 2017).

Administration Time Issues

Although these assessment-related issues played a role in shaping the content of these KEAs, a frequently mentioned precursor to content revision was related to teachers' time. For example, in Delaware, where the KEA is a customized version of the observational Teaching Strategies GOLD, the initial 2013 measure had 43 items. However, teacher feedback indicated that it took too much time to collect evidence for all 43 items within the allotted administration period. Consequently, the total number of items was reduced to 34 (Delaware Department of Education [DDOE], 2016d; ELCTA, 2014). Pennsylvania also removed more than half of the items from the 2011 pilot version of the state's KEA because of teacher concerns about the length of the measure (Pennsylvania OCDEL, 2014a, 2014b).

The state of Washington also uses a customized version of GOLD known as the Washington Kindergarten Inventory of Developing Skills (WaKIDS). On the basis of feedback from pilot teachers about their overall assessment workload and the 23 hours needed on average to complete this KEA (Build Initiative, 2012; Butts, 2013, 2014), and in light of the emergence of new K–12 learning standards, the item content of the WaKIDS was revised in 2015. Eight items related to social–emotional, language, and cognitive skills were added, but 13 previously administered items were removed, resulting in a total of 31, rather than 36, items (Office of the Governor, State of Washington, 2016; Washington Office of Superintendent of Public Instruction [Washington OSPI], 2014, 2015; Weisenfeld, 2017b). Teachers reported that this item decrease resulted in an average of 18 hours spent completing the observation (Butts, 2014).

Similarly, Maryland and Ohio teachers provided negative feedback on the amount of time necessary to administer the 2014 inaugural KEA, with reports of spending between 1 and 2 hours per student. To address this issue, in 2015, the measure's assessment developers removed 13 items that were more difficult or time intensive for teachers to administer but not considered critical to determining students' kindergarten readiness. This reduction resulted in a total of 50 items across four domains (Maryland State Department of Education [MSDE], 2015a, 2015b, 2016b, 2017c; Maryland State Education Association [MSEA], 2014; ODE, 2015; Office of Early Learning and School Readiness, Ohio Department of Education [ODE], 2016; Office of the Governor, State of Maryland, 2016; Office of the Governor, State of Ohio, 2016; Schachter, Strang, & Piasta, 2015, 2017; WestEd, 2014, 2015; Wiggins, 2014). In addition, based on the results of item-level analysis, a few items were moved to other domains (MSDE, 2016a, 2016b).

Issues Resulting in the Revision of Kindergarten Entry Assessment Timeline Policies

A second focus of my study was the on-the-ground issues that precipitated changes to policies regarding the timeline for administration of these KEAs. As measures that aim to focus on the skills and knowledge possessed by entering kindergartners, this specific policy is important if policy makers seek to reliably compare KEA data from classrooms across the state. As mentioned, states generally require KEAs to be administered within 30–60 days after the start of school (ELCTA, 2016). This variation is reflected in the current policies for the study's KEA sample as well. As is displayed in Table 3, my analysis of the document data collected as part of the current study suggests that the administration timelines for all seven KEAs investigated were modified in response to both time crunch and reliability issues.

Table 3 Issues Contributing to Kindergarten Entry Assessment Timeline Policy Revisions

Kindergarten entry assessment	Teacher time crunch	Equivalent administration windows
Delaware Early Learner Survey	X	
Illinois Kindergarten Individual Development Survey (KIDS)	X	
Maryland/Ohio Kindergarten Readiness Assessment	X	
North Carolina Kindergarten Entry Assessment	X	
Oregon Kindergarten Assessment		X
Pennsylvania Kindergarten Entry Inventory		X
Washington Kindergarten Inventory of Developing Skills (WaKIDS)	X	

Time Crunch Issues

As mentioned in the previous section, after the 2013 field test of Delaware's observational Early Learner Survey, teacher feedback indicated that it took too much time to collect evidence for all of the 43 original items within the allotted administration period. As a result, the total number of items was reduced to 34 (DDOE, 2016d; ELCTA, 2014). However, during the 2015 state rollout of the measure, some kindergarten teachers also found that it was difficult to collect evidence, use the scoring rubric, and enter score data in the 30-day window originally allocated to complete these tasks. In response, prior to the fall 2016 administration, policy makers added 15 additional days for data entry, as opposed to requiring teachers to both collect evidence and input scores within the first 30 days (DDOE, 2016c; Delaware Office of Early Learning, 2016; Office of the Governor, State of Delaware, 2016).

Time crunch issues also resulted in revisions to which KEA items were administered at specific points in the school year. In Illinois, teachers were originally asked to administer all 55 KIDS items within the first 40 days of school, halfway through the kindergarten year, and then again in late spring. However, teachers found it challenging to comply with this policy because of the amount of time required. Therefore the state then asked teachers to focus on items in five of the domains in the fall and spring and then to assess students on the remaining items during the mid-year administration only. The results of a teacher survey informed which items and domains were used during each of the assessment's three periods (Kriener-Althen & Hernandez, 2015; Office of the Governor, State of Illinois, 2016).

Illinois updated these requirements for the 2017–2018 school year. Currently teachers must complete 14 specific "kindergarten readiness" KIDS items within the first 40 days of school. These items are drawn from the Approaches to Learning—Self-Regulation, Social and Emotional Development, Language and Literacy Development, and Cognition: Math domains and were selected after reviewing the school readiness literature and conducting analyses aimed at discerning which items consistently demonstrated reliable scores. Teachers, schools, and districts also have the option of choosing to complete the full set of items within any of the measure's 11 domains (Illinois State Board of Education, 2017; Office of the Governor, State of Illinois, 2017; WestEd Center for Child and Family Studies, 2017; WestEd Center for Child and Family Studies, & University of California, Berkeley Evaluation and Assessment Research Center, 2017).

In North Carolina, 63% of the state's 2014 KEA pilot teachers indicated that collecting evidence for all 10 construct progressions was too time consuming when they also needed to learn how to use the measure's electronic data platform. This platform is critical, as it is used to house teachers' notes, photographs, and videos of a child demonstrating a particular item-relevant skill (e.g., naming letters of the alphabet). Therefore, in the beginning of the 2015–2016 school year, the Office of Early Learning asked teachers to administer just three of the construct progressions (book orientation, print awareness, and object counting). This number was increased to 7 of the 10 constructs in the 2016–2017 school year. During the 2017–2018 school year, kindergarten teachers were required to use 8 of the 10 constructs (Ferrara & Lambert, 2015, 2016; Ferrara, Merrill, Lambert, & Baddouh, 2017; North Carolina Office of Early Learning, 2015a, 2015b; Office of the Governor, State of North Carolina, 2015, 2016; Pruette, 2016, 2017).

Washington's KEA provides an example of a different type of administration timeline change in response to teacher feedback. More specifically, in addition to an observational KEA, WaKIDS includes a Family Connection component that involves a face-to-face meeting with parents as a means for helping teachers understand students' needs and backgrounds. During the 2012 WaKIDS pilot, teachers raised concerns about the amount of time necessary to schedule and conduct these meetings. As a result, in 2013, the legislature passed a bill giving teachers up to 3 days to complete this task. The time spent in these meetings can be counted as instructional hours as well (Butts, 2014; Dorn, 2013, 2016; Joseph, Cevasco, Lee, & Stull, 2010; Office of the Governor, State of Washington, 2013, 2014; Taylor, 2013).

Maryland and Ohio use the same Kindergarten Readiness Assessment but took different approaches to rectifying administration timeline issues. Prior to the 2014 inaugural administration of this measure, to ensure that teachers had sufficient time to administer the direct items and collect evidence for the observational items that are part of this KEA, Ohio's state legislature amended the end of the fall kindergarten assessment window from October 1 to November 1 (Office of the Governor, State of Ohio, 2014, 2015). Maryland's KEA assessment window originally ended in early November (MSDE Division of Early Childhood Development, 2015; Office of the Governor, State of Maryland, 2016; WestEd, 2014). However, as a means for further reducing administration time, in 2016, the state assembly passed legislation that requires only a random sample of kindergartners from each class to be assessed rather than all students. The end of the administration window was moved to early October as well. Individual schools and county boards of education still have the option of assessing all kindergartners with the measure (MSDE, 2017a, 2017c; MSEA, 2016; Weisenfeld, 2017b).

Ensuring Equivalent Administration Windows

The administration timelines for Pennsylvania's and Oregon's respective KEAs also were revised. However, in both of these cases, the revisions were aimed at reducing variations in when kindergarten teachers were assessing children and thus making it difficult to reliably compare scores across the state. In 2012, Pennsylvania moved the reporting deadline from November 15 to October 19 (Pennsylvania OCDEL, 2014a, 2014b). Currently teachers who use the state's KEA must complete the assessment within the first 45 calendar days of school (Pennsylvania OCDEL, 2015). In Oregon, because teachers had been assessing kindergartners from mid-August through late October, beginning in the 2013–2014 school year, the assessment window was narrowed to each district's first 6 weeks of school (Furrer & Green, 2013).

Supporting Teachers' Kindergarten Entry Assessment Administration and Observation Capacity

Regardless of what timeline is used, if kindergarten teachers who are tasked as KEA assessors or observers are to use that evidence to inform their instruction, they also need the capacity to produce reliable data. Therefore a third focus of my study was the issues that prompted modifications to the training and technical assistance provided to teachers related to administering a measure or collecting evidence via an observational rubric. As is displayed in Table 4, analyses of the study's data suggest that six of the seven sample KEAs needed to tweak the training or support related to the overall administration or observation process as well as accommodating students with special needs or who are considered to be English language learners and using technology platforms to upload data.

Overall Training and Supports

The rollout of Delaware's Early Learner Survey provides an example of some training and teacher support challenges that impacted changes to policy and practice. As was mentioned, this KEA is a customized version of Teaching Strategies GOLD. However, kindergarten teachers participating in the state's 2013–2014 field test indicated that they were confused by having to use the instructional manual for the full GOLD measure rather than their state's customized Early Learner Survey. In addition, teachers noted that the predetermined "learning kits" that were provided to help facilitate administration of this measure were not always aligned with the materials they actually needed. Then, during the 2015 full state rollout, teachers expressed frustration with the availability of online and in-person trainings. Teachers also reported that it was too challenging and time consuming to pass the online interrater reliability certification process, particularly because they lacked adequate support to help improve their scoring. Finally, teachers expressed the need for substitutes so that they could focus on inputting their evidence into the online platform (DDOE, 2016c; ELCTA, 2014; Office of the Governor, State of Delaware, 2016).

In response, Delaware's Office of Early Learning and the observation's developer collaborated to create a state-specific manual and resource guide that is available in both hard copy and online. All kindergarten teachers were given a $50 stipend to purchase customized classroom learning materials to be used during their administration of the KEA as well. In addition, districts could apply for a $100 reimbursement from the state to hire a one-day substitute to provide kindergarten teachers with adequate time to input all of their KEA data.[2] Furthermore, Delaware's KEA training policy was amended to provide 13 in-person or online training opportunities to take place during previously scheduled teacher professional development days as well as at locations convenient to teachers. The state also eliminated the online interrater

Table 4 Assessor/Observer Training and Support Revisions

Kindergarten entry assessment	Overall capacity	Testing accommodations	Technology
Delaware Early Learner Survey	X	X	X
Illinois Kindergarten Individual Development Survey (KIDS)			
Maryland/Ohio Kindergarten Readiness Assessment		X	X
North Carolina Kindergarten Entry Assessment	X		X
Oregon Kindergarten Assessment		X	
Pennsylvania Kindergarten Entry Inventory	X		X
Washington Kindergarten Inventory of Developing Skills (WaKIDS)	X		

certification requirement and instead offered the process to new administrators of the measure during their in-person trainings (DDOE, 2016c, 2016d; Delaware Office of Early Learning, 2016; Office of the Governor, State of Delaware, 2016).

During the 2014 pilot of North Carolina's observational KEA, some teachers reported difficulty in identifying evidence of children's skills as measured by the KEA's observation rubric. The majority of these pilot teachers reported that they needed more training on using the electronic data platform to record their observation scores as well. To rectify these issues, in 2015, the state's Office of Early Learning revised the manual that outlines all of the constructs assessed by the KEA and provides examples of classroom situations demonstrating the different levels of development within each construct. In addition, the office worked with the technology platform supplier to update it and provide an implementation guide (Ferrara & Lambert, 2015; North Carolina Office of Early Learning, 2015a, 2015b, 2015c, 2017b). More recent teacher feedback informed a redesign of the technology platform in terms of the look of the landing page, access to key pages, suitability for tablet devices, and filtering options when generating score reports (Public Schools of North Carolina, 2017).

In Pennsylvania, the issue was not the quality of the Kindergarten Entry Inventory training itself but instead the availability of teachers to participate in it. During the 2014 state rollout period, one key issue was that teachers in some districts were contractually prevented from attending the in-person training offered because it did not occur on a dedicated professional development day (Commonwealth of Pennsylvania Governor's Office, 2015). To address this issue, the state updated the training provided to include both face-to-face and Internet-based opportunities (Commonwealth of Pennsylvania Governor's Office, 2016; Pennsylvania Department of Education, 2016).

Washington provides an example of a different type of support that was provided to teachers as a means for increasing their administration/observation capacity. As was noted, teachers implementing the WaKIDS observational KEA experienced time crunch issues, and one response was to reduce the number of items. However, through a collaboration with the Bill and Melinda Gates Foundation, in fall 2013, the state also distributed noncompetitive grants to districts to pay for an array of WaKIDS administration supports. These supports included hiring substitutes so teachers could collect evidence for the WaKIDs and paying for a paraprofessional to enter WaKIDS data into the online platform. In 2014, the WaKIDS workgroup also recommended that districts not require kindergarten teachers to administer any other assessments during the assessment window (Butts, 2013, 2014; Dorn, 2013).

In addition, the interrater reliability process for WaKIDS underwent revision subsequent to research on this topic. During the fall 2012 pilot, achieving interrater reliability was optional, and only 174 out of 1,003 teachers completed the process in place at that time (Office of the Governor, State of Washington, 2013). Furthermore, analyses of data from a small fall 2012 study showed that interrater reliability was "moderate" overall, with at least 81% agreement within the Social, Physical, and Language domains; 75% for Literacy and Mathematics; and 68% for the Cognitive domain items. Teachers had the most difficulty correctly rating lower performing students and the language abilities of English language learners (Soderberg et al., 2013).

In response to these data, beginning in 2013, Washington's Office of Superintendent of Public Instruction made achieving rater reliability certification part of the training and offered financial incentives for teachers to complete the certification process (Office of the Governor, State of Washington, 2015; Washington OSPI, 2016). By the fall 2013 WaKIDS administration, 67% of trained teachers had earned their rater certification (Office of the Governor, State of Washington, 2014). This number grew to 83% of trained teachers by the fall 2014 (Office of the Governor, State of Washington, 2015) and fall 2015 administrations (Office of the Governor, State of Washington, 2016).

Accommodations Support

A related issue is the capacity to administer an assessment or collect evidence for an observational rubric when assessing students who are classified as having special needs or being English language learners. During Delaware's 2015 full state rollout, both teachers and administrators indicated the need for additional guidance about including kindergartners with disabilities as well as which students should be exempted from being observed. Teachers also reported the need for more support on using the state's KEA with English language learners. To address these concerns, the Department of Education added more specific guidelines for observing both populations of students with the Early Learning Survey. These guidelines were uploaded to the Early Learner Survey Web site as well (DDOE, 2016b, 2016c, 2016e, 2016f; Office of the Governor, State of Delaware, 2016).

Teachers administering Maryland and Ohio's Kindergarten Readiness Assessment can participate in a 2-day face-to-face, online, or blended training and have access to online resources, including supports for administering each KEA item

and a set of videos clips to practice score the measure. Professional development modules are available in teachers' online "KReady" assessment user support accounts as well. In addition, teachers are required to demonstrate their scoring reliability by successfully passing a simulated KEA and a 20-item, multiple choice content assessment (MSDE, 2015b, 2017c; MSDE Division of Early Childhood Development, 2015; MSEA, 2014; Office of the Governor, State of Maryland, 2015, 2016; WestEd, 2014). Yet, subsequent to the 2014 inaugural administration of this KEA, a survey conducted by the MSEA (2014) suggested that some teachers felt they did not have adequate instructions or resources to appropriately accommodate English language learner students. Additional results suggested that some teachers perceived that the measure was too difficult to administer to special needs students because of guidelines prohibiting certain accommodations. Although it is not clear if the survey's results played a direct role, by the fall 2015 administration of the measure, the assessment developers had revised the training modules on allowances and supporting individual children and updated the document that outlines the guidelines for allowable supports (MSDE, 2015b; WestEd, 2015).

Research conducted in Oregon during the 2012 pilot revealed that Oregon kindergarten teachers varied in their procedures for assessing English language learners. There also was a need for guidelines on appropriate accommodations for special needs students. All of this research suggested that the webinar trainings offered at that time did not appear to be sufficient for ensuring reliable administration of the different components of the state's KEA. Furthermore, after the statewide 2013–2014 field test, teachers indicated the need for further guidance on administering the assessment with Spanish-speaking English language learners (Early Learning Division, Oregon Department of Education, 2014, 2016; Furrer & Green, 2013).

In response, the Oregon Kindergarten Assessment was included in the state's Department of Education Accommodations Manual beginning in 2013–2014. And, beginning in 2014, the Oregon Department of Education (2015a) provided guidance via its "Decision Matrix" on identifying Spanish-speaking English language learners and on locating Spanish bilingual assessors to administer the literacy portion of the KEA. One additional clarification made was that kindergartners being assessed with the Spanish–English bilingual version of the Early Math subtest were allowed to provide an answer in either language (Early Learning Division, Oregon Department of Education, 2015, 2016; Furrer & Green, 2013; Office of Assessment and Accountability, Oregon Department of Education, 2016a; Oregon Department of Education, 2014).

Technology Issues

Delaware also provides an example of the impact of technology issues in revisions to the support provided to kindergarten teachers. For example, during the 2015 full state rollout, the online technology platform used to record children's Early Learner Survey scores was not user-ready, including teachers discovering that incorrect classroom rosters had been uploaded. In addition, the platform was not consistently available during the 30-day observation and data entry window (Office of the Governor, State of Delaware, 2016). To rectify these issues, the state incorporated new procedures to ensure that correct student rosters would be preloaded into the online technology platform. In addition, the assessment developer redesigned its help desk and provided a larger number of state-dedicated technical support staff. The state also identified experienced kindergarten teachers who could provide technical assistance to new platform users (DDOE, 2016c; Delaware Office of Early Learning, 2016).

Similarly, Maryland and Ohio kindergarten teachers involved in the 2014 inaugural administration of these states' Kindergarten Readiness Assessment reported various technology-related issues, including site crashes, incorrect student name spellings and birth dates, loss of entered data, and issues with children knowing how to use iPads (MSEA, 2014; WestEd, 2014). In response, the technology team improved the online system and designated dedicated technology trainers to assist teachers. In addition, prior to rolling out the improved system, the technology team tested it with 25 teachers and 4 data managers from both states. This enabled the team to address critical system bugs or issues before the actual cross-state launch (MSDE, 2015b; ODE, 2015b).

Pennsylvania teachers also experienced data entry system challenges during the 2014/Cohort I state rollout of its KEA owing to the overload caused by the sheer number of teachers using the system on their dedicated professional development days. To address these issues, the Office of Child Development and Early Learning created a new data system for the 2015 Kindergarten Entry Inventory data that was not dependent on a specific browser software, allowed users to connect via a tablet or smartphone, and reduced the amount of time between data entry and the system's response (Commonwealth of Pennsylvania Governor's Office, 2015, 2016; Pennsylvania Department of Education, 2015, 2016).

Finally, I could not find evidence that Illinois needed to amend the training and technical assistance provided to teachers related to collecting KIDS data or recording these data in the measure's online system. Each teacher is provided with 2 days of online module training, with the format and content based on resources and materials used to train teachers tasked with implementing the DRDP in California (and on which KIDS is based). Teachers also have online access to tutorials, resources, checklists, and the rater reliability system. In addition, teachers who participated in the pilot and field testing of the measure from 2013 to 2015 had access to regional coaches, who were financially supported through a $1.2 million grant from a Chicago-based foundation. The entire training and technical assistance system was developed, piloted, and completed in 2015 (Office of the Governor, State of Illinois, 2016, 2017).

Supporting Teachers' Access to and Use of Posttest Data

A final area of interest in my study was any training and technical support revisions that stemmed from teachers' self-reported capacity to access KEA data or use these data to inform their instruction. As noted, use of such data is one of the two primary premises for assessing entering kindergartners. Analysis of the study's data suggests that five of the study's seven KEAs needed to respond to access or interpretation issues (see Table 5).

Access to Data

Documents related to two KEAs contained information about data access issues that were subsequently addressed. During the 2015 full state rollout of Delaware's Early Learner Survey, the score results were not available at the level of detail that teachers or district stakeholders necessarily found to be useful (DDOE, 2016c). Although the documents I reviewed did not indicate why this was the case, in 2015, the state also experienced the technology platform issues that were highlighted earlier (Office of the Governor, State of Delaware, 2016). No matter the reason, by fall of the 2016–2017 school year, training was provided on how to run and use classroom- and school-level data. Furthermore, classroom reports could be run immediately after inputting any KEA results, and school- and district-level data were slated to be available in December (DDOE, 2016c; Delaware Office of Early Learning, 2016).

The first cohort of teachers who piloted Pennsylvania's KEA also reported issues with accessing relevant data from the state's system. The state subsequently enhanced the system and, in so doing, provided teachers with the capacity to print reports at both the child and classroom levels. The system also can generate item-level reports at the school, district, and state levels. In addition, directions for accessing the reports are available on the Kindergarten Entry Inventory landing page (Commonwealth of Pennsylvania Governor's Office, 2016).

Using Kindergarten Entry Assessment Data

For three additional KEAs, corrections stemmed from research demonstrating that teachers needing further guidance on using KEA data to inform their instruction. In Illinois, the state hired coaches to work directly with teachers after receiving feedback about their lack of capacity in this area (ELCTA, 2015a). During Washington's pilot of the WaKIDS assessment, some teachers indicated that the measure provided limited information to inform their instruction. The state responded to this issue by providing teachers with more information about the ways in which these data could be used (Butts, 2013, 2014).

Table 5 Accessing and Using Kindergarten Entry Assessment Data

Kindergarten entry assessment	Accessing data	Using data
Delaware Early Learner Survey	X	
Illinois Kindergarten Individual Development Survey (KIDS)		X
Maryland/Ohio Kindergarten Readiness Assessment		
North Carolina Kindergarten Entry Assessment		X
Oregon Kindergarten Assessment		
Pennsylvania Kindergarten Entry Inventory	X	
Washington Kindergarten Inventory of Developing Skills (WaKIDS)		X

In North Carolina, 57% of surveyed pilot KEA teachers reported that they struggled with figuring out how to use the measure's data as part of their classroom practice (Ferrara & Lambert, 2015). In response, the state's Office of Early Learning released a guidebook on interpreting the data and using it as evidence to make instructional decisions (North Carolina Office of Early Learning, 2015a). In addition, the office also offers a series of online, self-paced modules and facilitate courses on formative assessment and data literacy.[3]

Finally, kindergarten teachers using the two remaining KEAs did not have immediate access to their classrooms' KEA data at one point. However, in contrast to the other KEAs, this situation stemmed from the additional analyses and reviews that were conducted to support utility and validity. In Maryland and Ohio, the delay occurred after the 2014 inaugural administration of the KEA used in both states but was due to some planned psychometric analyses and a February 2015 meeting to determine performance standards. By fall 2015, teachers in both states were able to download individual and classroom-level raw data as soon as scores were entered into the online platform. These results—which are provided through what is known as a dashboard—also provide teachers with guidance about instructional activities that are appropriate for individual students and groups of children based on their scores (MSDE, 2015b, 2016b; ODE, 2015b; Office of the Governor, State of Maryland, 2016; Office of the Governor, State of Ohio, 2016; WestEd, 2015). Similarly, after the 2013–2014 statewide field test of Oregon's KEA, the data were not released until winter 2014 (Irvin et al., 2017). This was due to the convening of a panel of K–3 educators, administrators, and researchers to review classroom, school, district, and state report prototypes (Office of the Governor, State of Oregon, 2015).

Discussion and Implications

In this study, I investigated some of the assessment- and teacher-related issues that challenged the validity and reliability of seven KEAs and, in turn, prompted adjustments to the content of these measures, the policies regarding when they are administered or implemented, and the training and technical support provided to teachers to assess or observe students and use score data to inform their instruction. One purpose for conducting the study was to expand the early childhood education field's understanding of the shaping role these validity and reliability issues can play. Another purpose was to highlight the importance of iterative research as a means for both uncovering these issues and informing the policies and practices that can impact KEA validity and reliability. The study's results related to both of these aims have implications for early childhood education policy makers and researchers.

Importance of Anticipating Kindergarten Entry Assessment Policy and Practice Revisions

As mentioned earlier, if a theory of action undergirds the use a test, researchers have argued that investigations of the test's validity and reliability for specific purposes should inform relevant educational policies (Bennett, 2011b; Bennett et al., 2011; Chalhoub-Deville, 2016). To better understand this study's implications for education policy makers, it is helpful to first revisit the sample of seven KEAs used. They include Delaware's and Washington's customized versions of a commercially available observational assessment that is widely used in state-funded pre-K programs and Illinois's, North Carolina's, and Pennsylvania's respective state-developed measures. Oregon's KEA consists of literacy and mathematics items from a commercially available direct measure as well as items from an existing observational measure aimed at assessing students' approaches to learning. The final KEA—used in Maryland and Ohio—is a newly developed measure that includes both observational and direct items.

Given that six of the seven KEAs grew out of an existing measure or assessments used in other settings, at first glance, one might anticipate that any validity and reliability kinks had already been worked out. Indeed, revising an existing assessment is typically viewed as less expensive and time consuming as compared to developing a new measure and its related training materials, data systems, and implementation processes (Picus, Adamson, Montague, & Owens, 2010). However, when it comes to avoiding validity and reliability issues, the results of the current study suggest that prior usage of an earlier version of a measure, or full usage of a measure in other settings, does not necessarily present an advantage.

For example—and as displayed in Table 6—all seven KEAs experienced revisions to their content. In three cases, these revisions were admittedly to be expected, as they stemmed from incomplete alignment with kindergarten learning standards, which was a key focus of the RTT-ELC competition. However, research on the psychometric properties of the assessment also contributed to content revisions. Another major contributor was the impact of administering a KEA

Table 6 Issues and Revisions Experienced Across Study Kindergarten Entry Assessments (KEAs)

KEAs	KEA content	Administration timeline policies	Teacher training or support for: Overall administration capacity, using accommodations, or technology	Teacher training or support for: Accessing KEA data or using data to inform instruction
Customized version of Teaching Strategies GOLD				
Delaware Early Learner Survey	X	X	X	X
Washington Kindergarten Inventory of Developing Skills (WaKIDS)	X	X	X	X
New measure				
Maryland/Ohio Kindergarten Readiness Assessment	X	X	X	
Enhancements of existing state-developed measures				
Illinois Kindergarten Individual Development Survey (KIDS)	X	X		X
North Carolina Kindergarten Entry Assessment	X	X	X	X
Pennsylvania Kindergarten Entry Inventory	X	X	X	X
Items from several existing measures				
Oregon Kindergarten Assessment	X	X	X	

on teachers' instructional time, with time crunch issues resulting in a reduction in the total number of items within several KEAs.

Similarly, teacher feedback indicated the need for revised timeline policies for administering all seven KEAs. In four of these cases, the modifications were in response to time crunch issues. For a fifth KEA — which is used by two states — one state expanded the timeline to ensure sufficient administration/observation time, but the other state reduced the timeline after adopting a random sampling strategy. For the final two KEAs, these modifications were aimed at reducing variations in when kindergarten teachers were assessing children and thus making it difficult to reliably compare scores across the state.

As can also be seen in Table 6, analyses of the study's document-based data suggests that the training and ongoing technical support provided to kindergarten teachers tasked with using all seven KEAs needed to be modified in some way. In six cases, these modifications were aimed at teachers' administration/observation capacity and stemmed from the availability of training, available training materials, and the online platform in which KEA data are recorded or uploaded. For five of the study's KEAs, modifications were needed to the training or assistance provided to teachers to improve their capacity to access score data or use these data to inform their instruction.

Typically, policy makers make compromises early in the assessment selection or development process and in response to such top-down inputs as purpose, setting, cost, and current resources (Snow & Van Hemel, 2008). Although the results of this study help explain the why behind some of the current policies and practices related to these seven KEAs, when considered as a whole, the first key implication of this study is that bottom-up, real world compromises will likely need to be made, especially as a measure is developed, piloted, field tested, or rolled out on a large-scale basis. This may particularly be the case when the majority of teachers are not experienced users of the specific approach or measure used. In short, it may not be so much a question of if there are validity and reliability concerns but instead what they are and how they might be mitigated. Furthermore, such compromises may involve reconsidering a KEA's content, administration timeline policies, and the training and technical assistance provided to teachers.

Importance of Validity and Reliability Research for Informing Policy and Practice

Given the array of measure- and teacher-related validity and reliability challenges across the seven sample KEAs, a second key implication of this study is the importance of research that highlights these issues. Such research is critical not only

for ensuring the technical quality of a measure but also for determining to what extent policy and practices need to be modified to attain the goals of a KEA policy in a valid and reliable way. Of particular importance are any "on-the-ground" assessor and observer tweaks needed to support the theory of action driving the measure's use.

Analysis of the documents used to collect data for the current study demonstrates that validity- and reliability-related research can take many forms. For example, assessment developers, state workgroups, and other researchers conducted psychometric analyses of KEA data gathered through pilot and field tests to determine to what extent inferences might be drawn about entering kindergartners' knowledge and skills. In addition, they used qualitative surveys, interviews, and focus groups as a means for gathering critical teacher feedback about the supports needed to effectively administer these measures and use the resulting data to inform instruction.

I could not identify a common research model across all seven KEAs for investigating both measure validity and reliability and policy and practice adequacy. This is likely due, in part, to the fact that I purposefully selected a varied sample of KEAs both in terms of the approaches (direct vs. observational) and the measures used. Yet, the sample KEAs also experienced similar issues and revisions. It therefore could be that the more salient research implication is the importance of engaging in a customized plan, use, review, and revise research model to generate data on what assessment programmatic inputs are—and are not—supporting KEA validity and reliability. Furthermore, this study's results suggest the value of conducting research on an ongoing basis as opposed to only focusing on initial content or administration issues.

Future Research

The results of this study have additional implications for future research to be conducted as a means for informing KEA policy and practice. One key topic to explore is the extent to which the validity and reliability of all state-mandated or recommended KEAs across the United States have been investigated. This is important because, similar to the current study's sample of KEAs, the measures being used include customized versions of commercially available measures, newly developed assessments, and state-developed tools (Weisenfeld, 2017a).

Of course, although policy makers and other education stakeholders may realize the value of conducting both initial and ongoing validity and reliability research, one key variable to consider is the cost of obtaining that information (Riley-Ayers et al., 2011). And, as Weiss (1979) noted almost 40 years ago, the availability of research funds plays a role in the degree to which social scientists become interested in policy issues. Indeed, although the current study did not focus on the dollar amount spent on KEA validity and reliability-related research, an evaluation of Washington's RTT-ELC award found that the successful scaling up of WaKIDS was directly due to the receipt of these funds (Schilder, 2015). Presuming that there are states or developers that have not engaged in a methodologically rigorous set of studies, another topic for future research is how much KEA validity and reliability studies can cost. In addition, now that the RTT-ELC contracts have ended, of particular interest is whether the lack of such funds in the short term impacts states' capacity to conduct both psychometric research and investigations related to the policy and practice tweaks needed to bolster assessors,' observers,' and data users' capacities.

It also could be useful to build on the current study by taking a closer look at the validity and reliability trade-offs that come with making on-the-ground-related compromises to assessment content, administration timelines, and teacher training and support. Such studies might also be helpful for generating "lessons learned" and thus inform other states' KEA development and implementation efforts. One potential trade-off topic is the extent to which time crunch–precipitated item reductions impact the usefulness of KEA data for informing teachers' instruction or district and state policy makers' achievement gap–related decisions. Another potential trade-off topic is observer scoring reliability. For example, analysis of fall 2015 and 2016 administrations of Delaware's Early Learner Survey suggests some inconsistencies in how teachers are rating entering kindergartners' mathematics skills (Hanover Research, 2017). At the same time, only new administrators of the measure are asked to attain interrater certification as part of their training, and such reliability is considered to be "good to go" for 3 years (DDOE, 2017). A question for future research is how much and what kind of kindergarten-teacher-as-observer training and certification is optimal if KEA data are to inform both teachers' instruction and state policy decisions.

Finally, in the current study, many of the modifications made to the content of the sample KEAs, their respective administration timelines, and the training and support provided to teachers emanated from teacher self-report. Given that a test itself is neither valid nor invalid, but instead the use of scores for a particular purpose and population must be validated, to further understand the impact of any policy and practice compromises, future research should examine

to what extent kindergarten teachers are actually using KEA data to inform their instruction. Of additional interest is whether such data use is correlated with decreasing achievement gaps.

Limitations of the Study

Although this study provides new insight into the origin of some KEA policy and practice differences, it has two main limitations that hinder its generalizability. First, my sample was composed of seven purposefully selected KEAs used in eight states. Because at least 40 states and the District of Columbia have incorporated policies on these measures, my sample may not be representative of other states' KEA experiences. This may especially be the case given that one condition of my sample was the receipt of RTT-ELC funds.

A second limitation of this study is the nature of the data I analyzed. I relied on publicly available documents, including RTT-ELC annual progress reports, technical reports, and journal articles, to specifically seek out data on issues relevant to the study's three focus areas. Although I often was able to triangulate these data by looking for multiple sources for any single issue and resolution, it is not clear to what extent the information reported in any of these documents was complete. It is possible that there were additional issues and resolutions that I could not include here because they were not reported in a publicly available document. Future research should therefore build on these initial findings as a means for increasing the field's understanding of the factors to keep in mind when relying on document-based data to investigate the implementation of newly developed KEAs or previously used KEAs in new settings.

Conclusion

This study expands the early childhood education field's understanding of the validity and reliability issues that can shape evolving KEA assessment policies and practices. Such an understanding is useful given the widespread adoption of KEAs across the United States over the past 5 years. Furthermore, although KEAs have the potential to inform teachers' practice and state policy maker efforts to close school readiness gaps, these measures need to be psychometrically strong if they are to provide sufficient evidence for these purposes. In addition, the kindergarten teachers who are tasked as assessors, observers, and data users need the capacity to produce reliable scores and use that evidence to inform their instruction. The results of the current study demonstrate that merely implementing a KEA policy may not necessarily accomplish either goal. Instead, policy makers, assessment developers, and researchers need to work collaboratively to conduct research designed to highlight potential challenges to KEA validity and reliability and, in turn, inform the policy and practice compromises that should be made to support the use of KEA data for these important purposes.

Acknowledgments

I gratefully acknowledge the feedback received from W. Steven Barnett, Matt Brunetti, Sarah Daily, Caitlin Gleason, Kathleen Kebbeler, Phillip Irvin, Kerry Kriener-Althen, Richard Lambert, Donna Spiker, and Zijia Li, who served as external reviewers for this report, and from Brent Bridgeman and Donald Powers, who served as ETS technical reviewers. The views expressed in this report are the author's alone.

Notes

1 The Spanish literacy measure was temporarily suspended during the 2017–2018 school year so that researchers could determine if this is the most appropriate measure to use for this purpose (Oregon Department of Education, 2017a, 2017b).
2 The substitute reimbursement pool was eliminated in the 2017–2018 school year because of state budget reductions (DDOE, 2017).
3 http://nck-3fap.ncdpi.wikispaces.net/home; https://center.ncsu.edu/ncfalcon/

References

Ackerman, D. J., Barnett, W. S., Hawkinson, L. E., Brown, K., & McGonigle, E. (2009). *Providing preschool education for all 4-year-olds: Lessons from six state journeys*. New Brunswick, NJ: National Institute for Early Education Research.

Ackerman, D. J., & Coley, R. (2012). *State preK assessment policies: Issues and status*. Princeton, NJ: ETS Policy Information Center.

Akers, L., Del Grosso, P., Atkins-Burnett, S., Boller, K., Carta, J., & Wasik, B. A. (2014). *Tailored teaching: Teachers' use of ongoing child assessment to individualize instruction* (Vol. 2). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Alvarez, L., Ananda, S., Walqui, A., Sato, E., & Rabinowitz, S. (2014). *Focusing formative assessment on the needs of English language learners*. San Francisco, CA: WestEd.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, D., Alonzo, J., Tindal, G., Farley, D., Irvin, P. S., Lai, C., . . . Wray, K. A. (2014). *Technical manual: easyCBM* (Technical Report no. 1408). Eugene, OR: Behavioral Research and Teaching.

Atkins-Burnett, S., Bandel, E., & Aikens, N. (2012). *Assessment tools for the language and literacy development of young dual language learners (DLLs)* (Research Brief No. 9). Chapel Hill: University of North Carolina, FPG Child Development Institute, CECER-DLL.

Ball, C., & Gettinger, M. (2009). Monitoring children's growth in early literacy skills: Effects of feedback on performance and classroom environments. *Education and Treatment of Children, 32*, 189 – 212.

Banerjee, R., & Luckner, J. L. (2013). Assessment practices and training needs of early childhood professionals. *Journal of Early Childhood Teacher Education, 34*, 231 – 248.

Basford, J., & Bath, C. (2014). Playing the assessment game: An English early childhood education perspective. *Early Years: An International Research Journal, 34*, 119 – 132.

Bennett, R. E. (2011a). *CBAL: Results from piloting innovative K – 12 assessments* (Research Report No. RR-11-23). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2011.tb02259.x

Bennett, R. E. (2011b). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*, 5 – 25.

Bennett, R. E., Kane, M., & Bridgeman, B. (2011). *Theory of action and validity argument in the context of through-course summative assessment*. Princeton, NJ: Educational Testing Service.

Bonner, S. M. (2013). Validity in classroom assessment: Purposes, properties, and principles. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 87 – 106). Thousand Oaks, CA: Sage.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27 – 40.

Bowman, B., Donovan, S., & Burns, M. S. (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy of Sciences.

Bradbury, A. (2014). Early childhood assessment: Observation, teacher "knowledge" and the production of attainment data in early years settings. *Comparative Education, 50*, 322 – 339.

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child Behavior Rating Scale*. Cambridge, MA: Abt Associates.

Bronson, M. B., Tivnan, T., & Seppanen, P. S. (1995). Relations between teacher and classroom activity variables and the classroom behaviors of preschool children in chapter 1 funded programs. *Journal of Applied Developmental Psychology, 16*, 253 – 282.

Build Initiative. (2012). *Kindergarten entry assessment (KEA) question and answer session: WaKIDS: Washington State's KEA process*. Retrieved from http://www.buildinitiative.org/WhatsNew/ViewArticle/tabid/96/ArticleId/676/Kindergarten-Entry-Assessment-KEA-Question-and-Answer-Session-WaKIDS-Washington-States-KEA-Process.aspx

Build Initiative. (2016). *Kindergarten entry assessment — KEA*. Retrieved from http://www.buildinitiative.org/TheIssues/EarlyLearning/StandardsAssessment/KEA.aspx

Butts, R. (2013). *Recommendations of the Washington Kindergarten Inventory of Developing Skills (WaKIDS) workgroup*. Olympia, WA: Office of Superintendent of Public Instruction Early Learning Office.

Butts, R. (2014). *Recommendations of the Washington Kindergarten Inventory of Developing Skills (WaKIDS) workgroup*. Olympia, WA: Office of Superintendent of Public Instruction Early Learning Office.

Cabell, S. Q., Justice, L. M., Zucker, T. A., & Kilday, C. R. (2009). Validity of teacher report for assessing the emergent literacy skills of at-risk preschoolers. *Language, Speech, and Hearing Services in Schools, 40*, 161 – 173.

California Department of Education Child Development Division. (2012). *Desired Results Developmental Profile: School Readiness (DRDP-SR)*. Sacramento, CA: Author.

California Department of Education Child Development Division. (2015). *Desired Results Developmental Profile: Kindergarten (DRDP-K)*. Sacramento, CA: Author.

Campbell, C. (2013). Research on teacher competency in classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 71 – 84). Thousand Oaks, CA: Sage.

Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on early identification of reading disabilities. *Journal of Learning Disabilities, 42*, 163 – 176.

Center on Standards and Assessment Implementation. (2017). *State of the states: Pre-K/K assessment*. Retrieved from http://www.csai-online.org/sos

Chalhoub-Deville, M. (2016). Validity theory: Reform policies, accountability testing, and consequences. *Language Testing, 33*, 453 – 472.

Commonwealth of Pennsylvania Governor's Office. (2015). *Race to the Top Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Commonwealth of Pennsylvania Governor's Office. (2016). *Race to the Top Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Connors-Tadros, L. (2014). *Information and resources on developing state policy on kindergarten entry assessment (KEA)* (CEELO Fast-Fact). New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *American Journal of Medicine, 119*, 166.e7 – 166.e16.

Cox, M. E., Rodriguez, M. C., & Edwards, K. (n.d.). Empirical alignment of assessments to standards: A new direction for kindergarten entry. Roseville, MN: Minnesota Department of Education. Retrieved from http://www.education.state.mn.us/mdeprod/groups/educ/documents/hiddencontent/bwrl/mdyw/~edisp/mde060047.pdf

Davy, T., Ferrara, S., Holland, P. W., Shavelson, R., Webb, N. M., & Wise, L. L. (2015). *Psychometric considerations for the next generation of performance assessment*. Princeton, NJ: Educational Testing Service.

Delaware Department of Education. (2016a). *2016 – 17 Delaware Early Learner Survey implementation window dates: School district schedule*. Retrieved from http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/432/2016-17%20DE-ELS%20District%20Implementation%20Window%20Dates.pdf

Delaware Department of Education. (2016b). *DE-ELS frequently asked questions 2016 – 2017*. Dover, DE: Author.

Delaware Department of Education. (2016c). *DE-ELS updates and improvements 2016 – 2017*. Dover, DE: Author.

Delaware Department of Education. (2016d). *Delaware Early Learner Survey 2016 – 2017 fact sheet*. Dover, DE: Author.

Delaware Department of Education. (2016e). *Guidance for surveying dual- and English-language learners*. Dover, DE: Author.

Delaware Department of Education. (2016f). *Guidelines for understanding least restrictive environments*. Dover, DE: Author.

Delaware Department of Education. (2017). *Delaware Early Learner Survey updates 2017 – 2018 school year*. Dover, DE: Author.

Delaware Office of Early Learning. (2016). *Delaware Early Learner Survey recorded orientation presentation*. Dover, DE: Delaware Department of Education.

Dever, M. T., & Barta, J. J. (2001). Standardized entrance assessment in kindergarten: A qualitative analysis of the experiences of teachers, administrators, and parents. *Journal of Research in Childhood Education, 15*, 220 – 233.

Division for Early Childhood of the Council for Exceptional Children. (2007). *Promoting positive outcomes for children with disabilities: Recommendations for curriculum, assessment, and program evaluation*. Missoula, MT: Author.

Division for Early Childhood of the Council for Exceptional Children. (2010). *Responsiveness to ALL children, families, and professionals: Integrating cultural and linguistic diversity into policy and practice* (DEC position statement). Missoula, MT: Author.

Dorn, R. I. (2013, August 22). *Assessment and student information/early learning education* (Memorandum No. 37-12). Olympia, WA: Office of the Superintendent of Public Instruction.

Dorn, R. I. (2016, April 11). *Assessment and student information/teaching and learning/early learning* (Bulletin No. 012-16). Olympia, WA: Office of the Superintendent of Public Instruction.

Early Learning Challenge Technical Assistance. (2014). *Engaging and supporting educators in the development and implementation of KEA*. Retrieved from https://elc.grads360.org/services/PDCService.svc/GetPDCDocumentFile?fileId=5738

Early Learning Challenge Technical Assistance. (2015a). *National working meeting on early learning assessment: Meeting summary*. Retrieved from https://elc.grads360.org/#program/national-working-meeting-on-early-learning-assessment

Early Learning Challenge Technical Assistance. (2015b). *Statewide KEA data collection and reporting in RTT-ELC states*. Retrieved from https://elc.grads360.org/services/PDCService.svc/GetPDCDocumentFile?fileId=14468

Early Learning Challenge Technical Assistance. (2016). *Kindergarten entry assessments in RTT-ELC grantee states*. Retrieved from https://elc.grads360.org/services/PDCService.svc/GetPDCDocumentFile?fileId=23510

Early Learning Division, Oregon Department of Education. (2014). *Race to the Top Early Learning Challenge 2013 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Early Learning Division, Oregon Department of Education. (2015). *Race to the Top Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Early Learning Division, Oregon Department of Education. (2016). *Race to the Top Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Early Learning Division, Oregon Department of Education. (2017). *Race to the Top Early Learning Challenge 2016 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Education Commission of the States. (2014). *Kindergarten entrance assessments: 50-state comparison*. Retrieved from http://ecs.force.com/mbdata/mbquestRT?rep=Kq1407

Feeney, S., & Freeman, N. K. (2014, March). Standardized testing in kindergarten. *Young Children,* pp. 84 – 88.

Ferrara, A. M., & Lambert, R. G. (2015). *Findings from the 2014 North Carolina kindergarten entry formative assessment pilot*. Charlotte, NC: Center for Educational Measurement and Evaluation, University of North Carolina at Charlotte.

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, NC: Center for Educational Measurement and Evaluation, University of North Carolina at Charlotte.

Ferrara, A. M., Merrill, E., Lambert, R. G., & Baddouh, P. G. (2017). Effects of district specific professional development on teacher perceptions of a statewide kindergarten formative assessment. Paper presented at the 2017 annual meeting of the American Educational Research Association, San Antonio, TX.

Furnari, E. C., Whittaker, J., Kinzie, M., & DeCoster, J. (2016). Factors associated with accuracy in prekindergarten teacher ratings of students' mathematical skills. *Journal of Psychoeducational Assessment* https://doi.org/10.1177/0734282916639195, *35*, 410, 423

Furrer, C., & Green, B. L. (2013). *Oregon Kindergarten Assessment fall 2012 pilot process evaluation*. Portland, OR: Center for Improvement of Child and Family Services, Portland State University.

Garon, N., Smith, I. M., & Bryson, S. E. (2014). A novel executive function battery for preschoolers: Sensitivity to age differences. *Child Neuropsychology, 20*, 713–736.

Gibton, D. (2016). *Researching education policy, public policy, and policymakers*. New York, NY: Routledge.

Gokiert, R., Noble, L., & Littlejohns, L. B. (2013). Directions for professional development: Increasing knowledge of early childhood measurement. *Dialog*, *16*(3), 1–20.

Golan, S., Wechsler, M., Cassidy, L., Chen, W.-B., Wahlstrom, K., & Kundin, D. (2013). *The McKnight Foundation Education and Learning Program PreK–Third Grade literacy and alignment formative evaluation findings*. Menlo Park, CA: SRI International.

Golan, S., Woodbridge, M., Davies-Mercier, B., & Pistorino, C. (2016). *Case studies of the early implementation of kindergarten entry assessments*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

Goldfeld, S., Sayers, M., Brinkman, S., Silburn, S., & Oberklaid, F. (2009). The process and policy challenges of adapting and implementing the Early Development Instrument in Australia. *Early Education and Development, 20*, 978–991.

Goldstein, J., & Flake, J. K. (2016). Towards a framework for the validation of early childhood assessment systems. *Educational Assessment, Evaluation and Accountability, 28*, 273–293.

Goldstein, J., & McCoach, D. B. (2011). The starting line: Developing a structure for teacher ratings of students' skills at kindergarten entry. *Early Childhood Research & Practice*, 13(2). Retrieved from http://ecrp.uiuc.edu/v13n2/goldstein.html

Goldstein, J., McCoach, D. B., & Yu, H. (2017). The predictive validity of kindergarten readiness judgments: Lessons from one state. *Journal of Educational Research, 110*(1), 50–60.

Gottfried, M. A., Stecher, B. M., Hoover, M., & Cross, A. B. (2011). *Federal and state roles and capacity for improving schools*. Santa Monica, CA: RAND Corporation.

Greenberg, J., & Walsh, K. (2012). *What teacher preparation programs teach about K–12 assessment: A review*. Washington, DC: National Council on Teacher Quality.

Grisham-Brown, J., Hallam, R. A., & Pretti-Frontczak, K. (2008). Preparing Head Start personnel to use a curriculum-based assessment. *Journal of Early Intervention, 30*, 271–281.

Hanover Research. (2017). *Delaware Early Learner Survey key findings*. Arlington, VA: Author.

Harvey, E. A., Fischer, C., Weieneth, J. L., Hurwitz, S. D., & Sayer, A. G. (2013). Predictors of discrepancies between informants' ratings of preschool-aged children's behavior: An examination of ethnicity, child characteristics, and family functioning. *Early Childhood Research Quarterly, 28*, 668–682.

Harvey, H., Ohle, K., & Leshan, S. (2017). *The Alaska Developmental Profile: Perceptions of a state-mandated kindergarten screening tool*. Paper presented at the annual meeting of the American Educational Research Association, San Antonio.

Heroman, C., Burts, D. C., Berke, K., & Bickart, T. (2010). *Teaching Strategies GOLD objectives for development and learning: Birth through kindergarten*. Washington, DC: Teaching Strategies.

Heywood, T. (2015). Assessment. In D. Matheson (Ed.), *Introduction to the study of education* (pp. 118–133). Milton Park, England: Routledge.

Howard, E. C., Dahlke, K., Tucker, N., Liu, F., Weinberg, E., Williams R., … Brumley, B. (2017). *Evidence-based Kindergarten Entry Inventory for the commonwealth: A journey of ongoing improvement*. Washington, DC: American Institutes for Research.

Illinois State Board of Education. (n.d.). *KIDS Kindergarten Individual Development Survey frequently asked questions (FAQ)*. Retrieved from https://www.illinoiskids.org/content/faq-0

Illinois State Board of Education. (2015). *Kindergarten Individual Development Survey*. Springfield, IL: Author.

Illinois State Board of Education. (2017). *An overview on measures and data reporting*. Springfield, IL: Author.

Illinois State Board of Education Division of Early Childhood Education. (n.d.). *Illinois early learning standards: Kindergarten*. Springfield, IL: Author.

Irvin, P. S., Saven, J. L., Alonzo, J., Park, B. J., Anderson, D., & Tindal, G. (2012). *The development and scaling of the easyCBM CCSS elementary mathematics measures: Grade K* (Technical Report No. 1314). Eugene, OR: Behavioral Research & Teaching, University of Oregon.

Irvin, P. S., Tindal, G., & Slater, S. (2017). *Examining the factor structure and measurement invariance of a large-scale kindergarten entry assessment*. Paper presented at the annual conference of the American Educational Research Association, San Antonio, TX.

Isaacs, J., Sandstrom, H., Rohacek, M., Lowenstein, C., Healy, O., & Gearing, M. (2015). *How Head Start grantees set and use school readiness goals* (OPRE Report No. 2015-12a). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Jester, E. (2014, September 9). Teacher takes a stand, refused to give standardized test. *Gainesville Sun*. Retrieved from http://www.gainesville.com/news/20140909/teacher-takes-a-stand-refuses-to-give-standardized-test

Jiban, C. (2013). *Early childhood assessment: Implementing effective practice—A research-based guide to inform assessment planning in the early grades*. Portland, OR: NWEA.

Johnson, T. G., & Buchanan, M. (2011). Constructing and resisting the development of a school readiness survey: The power of participatory research. *Early Childhood Research & Practice, 13*(1). Retrieved from http://ecrp.uiuc.edu/v13n1/johnson.html

Jonas, D. L., & Kassner, L. (2014). *Virginia's Smart Beginnings kindergarten readiness assessment pilot: Report from the Smart Beginnings 2013/14 school year pilot of Teaching Strategies GOLD in 14 Virginia school divisions*. Richmond, VA: Virginia Early Childhood Foundation.

Jones, M., & Vickers, D. (2011). *Considerations for performance scoring when designing and developing next generation assessments* (White paper). Retrieved from http://images.pearsonassessments.com/images/tmrs/Performance_Scoring_for_Next_Gen_Assessments.pdf

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*, 130–144.

Joseph, G. E., Cevasco, M., Lee, T., & Stull, S. (2010). *WaKIDS pilot: Preliminary report*. Seattle, WA: Childcare Quality & Early Learning Center for Research and Training, University of Washington.

Kallemeyn, L. M., & DeStefano, L. (2009). The (limited) use of local-level assessment system: A case study of the Head Start National Reporting System and on-going child assessments in a local program. *Early Childhood Research Quarterly, 24*, 157–174.

Kamler, C., Moiduddin, E., & Malone, L. (2014). *Using multiple child assessments to inform practice in early childhood programs: Lessons from Milpitas Unified School District*. Princeton, NJ: Mathematica Policy Research.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73.

Karelitz, T. M., Parrish, D. M., Yamada, H., & Wilson, M. (2010). Articulating assessments across childhood: The cross-age validity of the Desired Results Developmental Profile-Revised. *Educational Assessment, 15*, 1–26.

Kilday, C. R., Kinzie, M. B., Mashburn, A. J., & Whittaker, J. V. (2012). Accuracy of teacher judgments of preschoolers' math skills. *Journal of Psychoeducational Assessment, 30*, 148–159.

Kim, K. (2016). Teaching to the data collection? (Un)intended consequences of online child assessment system, "Teaching Strategies GOLD". *Global Studies of Childhood, 6*(1), 98–112.

Kriener-Althen, K., Burmester, K., & Mangione, P. (2015). *Bridging early learning, kindergarten, and first-grade content standards through assessment*. Paper presented at the annual meeting of the American Education Research Association, Chicago, IL.

Kriener-Althen, K., & Hernandez, R. (2015). *KIDS school readiness brief: KIDS administration schedule informed by teacher input*. Paper presented at the annual meeting of the American Education Research Association, Chicago, IL.

Lai, C., Alonzo, J., & Tindal, G. (2013). *EasyCBM reading criterion related validity evidence: Grades K–1*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Lambert, R. G. (2003). Considering purpose and intended use when making evaluations of assessment: A response to Dickinson. *Educational Researcher, 32*(4), 23–26.

Lane, S. (2013). Performance assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 313–329). Thousand Oaks, CA: Sage.

Loesch-Griffin, D., Christiansen, E., Everts, J., Englund, L., & Ferrara, M. (2014). *Silver State Kindergarten Inventory Development Statewide (SSKIDS) pilot evaluation: Findings from Nevada's users of the Teaching Strategies GOLD (TSG) assessment tool*. Virginia City, NV: Turning Point.

Lopez, A. A., Turkan, S., & Guzman-Orth, D. (2017). *Conceptualizing the use of translanguaging in initial content assessments for newly arrived emergent bilingual students* (Research Report No. RR-17-07). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12140

MacDonald, M. (2007). Toward formative assessment: The use of pedagogical documentation in early elementary classrooms. *Early Childhood Research Quarterly, 22*, 232–242.

Martineau, J., Domaleski, C., Egan, K., Patelis, T., & Dadey, N. (2015). *Recommendations for addressing the impact of test administration interruptions and irregularities*. Washington, DC: Council of Chief State School Officers.

Maryland State Department of Education. (2015a). *Readiness matters! The 2014–2015 Kindergarten Readiness Assessment report*. Baltimore, MD: Author.

Maryland State Department of Education. (2015b). *Report on the Kindergarten Readiness Assessment (KRA): Joint chairmen's report (HB70)*. Baltimore, MD: Author.

Maryland State Department of Education. (2016a). *Readiness matters: Informing the future*. Baltimore, MD: Author.

Maryland State Department of Education. (2016b). *Readiness matters! The 2015–2016 Kindergarten Readiness Assessment report*. Baltimore, MD: Author.

Maryland State Department of Education. (2017a). *Readiness matters: Informing the future*. Baltimore, MD: Author.

Maryland State Department of Education. (2017b). *Student testing calendar school year 2017–2018*. Baltimore, MD: Author.

Maryland State Department of Education. (2017c). *The 2016–2017 Kindergarten Readiness Assessment technical report*. Baltimore, MD: Author.

Maryland State Department of Education Division of Early Childhood Development. (2015). *Kindergarten Readiness Assessment administration guide version 1*.5. Baltimore, MD: Author.

Maryland State Education Association. (2014). *MSEA report and recommendations on the Kindergarten Readiness Assessment*. Annapolis, MD: Author.

Maryland State Education Association. (2016). *Maryland General Assembly changes kindergarten readiness assessment to a sampling test*. Retrieved from http://www.marylandeducators.org/press/maryland-general-assembly-changes-kindergarten-readiness-assessment-sampling-test

Mashburn, A. J., & Henry, G. T. (2005). Assessing school readiness: Validity and bias in preschool and kindergarten teachers' ratings. *Educational Measurement: Issues and Practice, 23*, 16–30.

McClellan, C., Atkinson, M., & Danielson, C. (2012). *Teacher evaluator training and certification: Lessons learned from the Measures of Effective Teaching Project*. San Francisco, CA: Teachscape.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). San Francisco, CA: Jossey-Bass.

Messick, S. (1987). *Validity* (Research Report No. RR-87-40). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/j.2330-8516.1987.tb00244.x

Miller-Bains, K. L., Russo, J. M., Williford, A. P., DeCoster, J., & Cottone, E. A. (2017). Examining the validity of a multidimensional performance-based assessment at kindergarten entry. *AERA Open, 3*(2), 1–16.

Mislevy, R. J., Wilson, M. R., Ercikan, K., & Chudowsky, N. (2003). Psychometric principles in student assessment. In D. Stufflebeam & T. Kellaghan (Eds.), *International handbook of educational evaluation* (pp. 489–532). Dordrecht, Netherlands: Kluwer Academic Press.

Muller, U., Kerns, K. A., & Konkin, K. (2012). Test–retest reliability and practice effects of executive function tasks in preschool children. *Clinical Neuropsychology, 26*, 271–287.

Najarian, M., Snow, K., Lennon, J., Kinsey, S., & Mulligan, G. (2010). *Early childhood longitudinal birth cohort (ECLS-B): Preschool–kindergarten 2007 psychometric report*. Washington, DC: National Center for Education Statistics, Institute for Education Sciences.

National Association for the Education of Young Children. (2009). *Where we stand on assessing young English language learners*. Washington, DC: Author.

National Association for the Education of Young Children & National Association of Early Childhood Specialists in State Departments of Education. (2003). *Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age 8* Washington, DC: NAEYC.

National Head Start Association. (2016). *Kindergarten entry assessments*. Retrieved from https://www.nhsa.org/kindergarten-entry-readiness-assessments

North Carolina K–3 Assessment Think Tank. (2013). *Assessment for learning and development in K–3*. Raleigh, NC: Author.

North Carolina Office of Early Learning. (2015a). *K–3 formative assessment: Kindergarten. District implementation teams: Facilitator's guide. Interpreting and using evidence to guide instruction. Engaging families in the K-3 formative assessment process*. Raleigh, NC: Author.

North Carolina Office of Early Learning. (2015b). *K–3 formative assessment: Kindergarten. District implementation teams: Facilitator's guide. Using the formative assessment process with the technology platform*. Raleigh, NC: Author.

North Carolina Office of Early Learning. (2015c). *NC construct progressions and situations*. Raleigh, NC: Author.

North Carolina Office of Early Learning. (2017a). *Construct progression comparison document*. Raleigh, NC: Author.

North Carolina Office of Early Learning. (2017b). *NC construct progressions and situations*. Raleigh, NC: Author.

Office of Assessment and Accountability, Oregon Department of Education. (2016a). *Oregon kindergarten assessment specifications 2016–2017*. Salem, OR: Author.

Office of Assessment and Accountability, Oregon Department of Education. (2016b). *Preliminary test administration manual 2016–2017 school year*. Salem, OR: Author.

Office of Early Learning and School Readiness, Ohio Department of Education. (2016). *Ohio annual report on the Kindergarten Readiness Assessment*. Columbus, OH: Author.

Office of Governor Pat Quinn. (2011). *State of Illinois Race to the Top—Early Learning Challenge application for initial funding: From birth to kindergarten and beyond*. Springfield, IL: Author.

Office of Teaching, Learning, and Assessment, Oregon Department of Education. (2017). *Oregon kindergarten assessment specifications 2017–18*. Salem, OR: Author.

Office of the Governor, State of California. (2016). *Race to the Top—Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Delaware. (2016). *Race to the Top—Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Illinois. (2016). *Race to the Top—Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Illinois. (2017). *Race to the Top—Early Learning Challenge 2016 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Maryland. (2014). *Race to the Top—Early Learning Challenge 2013 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Maryland. (2015). *Race to the Top—Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Maryland. (2016). *Race to the Top—Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of North Carolina. (2013). *Race to the Top—Early Learning Challenge 2012 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of North Carolina. (2014). *Race to the Top—Early Learning Challenge 2013 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of North Carolina. (2015). *Race to the Top—Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of North Carolina. (2016). *Race to the Top—Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Ohio. (2014). *Race to the Top Early Learning Challenge 2013 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Ohio. (2015). *Race to the Top Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Ohio. (2016). *Race to the Top Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Washington. (2013). *Race to the Top Early Learning Challenge 2012 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Washington. (2014). *Race to the Top Early Learning Challenge 2013 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Washington. (2015). *Race to the Top Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Washington. (2016). *Race to the Top Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Ohio Department of Education. (2015). *Update on Kindergarten Readiness Assessment reductions*. Columbus, OH: Author.

Ohio Department of Education (2017). *Ohio's Kindergarten Readiness Assessment*. Retrieved from http://education.ohio.gov/Topics/Early-Learning/Kindergarten/Ohios-Kindergarten-Readiness-Assessment

Ohio Department of Education Office of Curriculum and Assessment. (2017). *Ohio's state tests rules book*. Columbus, OH: Author.

Oregon Department of Education. (2014). *Guidance for locating a Spanish bilingual assessor for the Oregon Kindergarten Assessment*. Retrieved from http://www.ode.state.or.us/news/announcements/announcement.aspx?=9909

Oregon Department of Education. (2015a). *Decision matrix for the early Spanish literacy assessment*. Salem, OR: Author.

Oregon Department of Education. (2015b). *Oregon Kindergarten Assessment report overview*. Salem, OR: Author.

Oregon Department of Education. (2017a). *2017–18 kindergarten assessment early literacy and early math training: Required for DTCs, STCs, and TAs* [Powerpoint presentation]. Salem, OR: Author.

Oregon Department of Education. (2017b). *Test administration manual: 2017–18 school year*. Salem, OR: Author.

Parkes, J. (2013). Reliability in classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 107–123). Thousand Oaks, CA: Sage.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

Pennsylvania Department of Education. (2015). *Pennsylvania Kindergarten Entry Inventory cohort 1 (2014) summary report*. Harrisburg, PA: Author.

Pennsylvania Department of Education. (2016). *Pennsylvania Kindergarten Entry Inventory cohort 2 (2015) summary report*. Harrisburg, PA: Author.

Pennsylvania Office of Child Development and Early Learning. (2013). *The 2011 SELMA pilot*. Harrisburg, PA: Commonwealth of Pennsylvania, Department of Education, Department of Public Welfare.

Pennsylvania Office of Child Development and Early Learning. (2014a). *The 2012 Pennsylvania Kindergarten Entry Inventory pilot*. Harrisburg, PA: Commonwealth of Pennsylvania, Department of Education, Department of Public Welfare.

Pennsylvania Office of Child Development and Early Learning. (2014b). *The 2013 Pennsylvania Kindergarten Entry Inventory pilot*. Harrisburg, PA: Commonwealth of Pennsylvania, Department of Education, Department of Public Welfare.

Pennsylvania Office of Child Development and Early Learning. (2015). *A guide to using the Pennsylvania Kindergarten Entry Inventory*. Harrisburg, PA: Author.

Pennsylvania Office of Child Development and Early Learning. (2017). *A guide to using the Pennsylvania Kindergarten Entry Inventory*. Harrisburg, PA: Author.

Picus, L. O., Adamson, F., Montague, W., & Owens, M. (2010). *A new conceptual framework for analyzing the costs of performance assessment*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Pruette, J. (2016, April 5). *KEA implementation, Phase II* (memorandum). Raleigh, NC: Office of Early Learning.

Pruette, J. (2017, March 16). *NC kindergarten entry assessment* (memorandum). Raleigh, NC: Office of Early Learning.

Public Schools of North Carolina. (2015). *District implementation teams: Facilitator's guide; Using the formative assessment process with the technology platform*. Raleigh, NC: Author.

Public Schools of North Carolina. (2017). *NC K–3 formative assessment process: KEA updates*. Retrieved from http://rtt-elc-k3assessment.ncdpi.wikispaces.net/file/view/1.%20CCSA%20KEA%20Updates%202017_FOR%20POSTING%20ON%20WIKI.pdf/610596145/1.%20CCSA%20KEA%20Updates%202017_FOR%20POSTING%20ON%20WIKI.pdf

Pyle, A., & DeLuca, C. (2017). Assessment in play-based kindergarten classrooms: An empirical study of teacher perspectives and practices. *Journal of Educational Research, 110*, 457–466.

Quirk, M., Nylund-Gibson, K., & Furlong, M. (2013). Exploring patterns of Latino/a children's school readiness at kindergarten entry and their relations with Grade 2 achievement. *Early Childhood Research Quarterly, 28*, 437–449.

Quirk, M., Rebelez, J., & Furlong, M. (2014). Exploring the dimensionality of a brief school readiness screener for use with Latino/a children. *Journal of Psychoeducational Assessment, 32*, 259–264.

Regenstein, E., Connors, M., Romero-Jurado, R., & Weiner, J. (2017). *Uses and misuses of kindergarten readiness assessment results* (Ounce Policy Conversations, Conversation No. 6, Version 1.0). Chicago, IL: Ounce of Prevention Fund.

Riley-Ayers, S. (2014). *Formative assessment: Guidance for early childhood policymakers* (CEELO policy report). New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.

Riley-Ayers, S., Frede, E., Barnett, S., & Brenneman, K. (2011). *Improving early education programs through data-based decision making*. New Brunswick, NJ: National Institute for Early Education Research.

Roach, A. T., McGrath, D., Wixson, C., & Talapatra, D. (2010). Aligning an early childhood assessment to state kindergarten content standards: Application of a nationally recognized alignment framework. *Educational Measurement: Issues and Practice, 29*, 25–37.

Robinson, J. P. (2010).The effects of test translation on young English learners' mathematical performance. *Educational Researcher, 39*, 582–590.

Rock, D. A., & Pollack, J. M. (2002). *Early childhood longitudinal study: Kindergarten class of 1998–99 (ECLS-K)—psychometric report for kindergarten through first grade*. Washington, DC: National Center for Education Statistics.

Roehrig, A. D., Duggar, S. W., Moats, L., Glover, M., & Mincey, B. (2008). When teachers work to use progress monitoring to inform literacy instruction. *Remedial and Special Education, 29*, 364–382.

Schachter, R. E., Strang, T. M., & Piasta, S. B. (2015). *Using the new Kindergarten Readiness Assessment: What do teachers and principals think?* Columbus, OH: Crane Center for Early Childhood Research and Policy, Ohio State University.

Schachter, R. E., Strang, T. M., & Piasta, S. B. (2017). Teachers' experiences with a state-mandated kindergarten readiness assessment. *Early Years*. https://doi.org/10.1080/09575146.2017.1297777

Schilder, D. (2015). *Washington State Race to the Top—Early Learning Challenge evaluation report*. Boston, MA: Build Initiative.

Schultz, T. (2014). *Top 9 reasons to implement a state KEA*. Paper presented at the NAEYC 2014 Professional Development Institute. Retrieved from http://ceelo.org/wp-content/uploads/2014/06/pdi_2014_kea_session_-9_reasons_handout.pdf

Scott-Little, C., Kagan, S. L., Reid, J. L., Sumrall, T. C., & Fox, E. A. (2014). *Common early learning and development standards analysis for the North Carolina EAG Consortium—Summary report*. Boston, MA: Build Initiative.

Shepard, L. A. (1997). Children not ready to learn? The invalidity of school readiness testing. *Psychology in the Schools, 34*, 85–97.

Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: National Academies Press.

Soderberg, J. S., Stull, S., Cummings, K., Nolen, E., McCutchen, D., & Joseph, G. (2013). *Inter-rater reliability and concurrent validity study of the Washington Kindergarten Inventory of Developing Skills (WaKIDS)*. Seattle, WA: University of Washington.

Stedron, J. M., & Berger, A. (2010). *NCSL technical report: State approaches to school readiness assessment*. Washington, DC: National Conference of State Legislatures.

Strand, P. S., Cerna, S., & Skucy, J. (2007). Assessment and decision-making in early childhood education and intervention. *Journal of Child and Family Studies, 16*, 209–218.

Strauss, V. (2014, September 15). Florida drops test after kindergarten teacher took public stand against it. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2014/09/15/florida-drops-test-after-kindergarten-teacher-took-public-stand-against-it/?utm_term=.2ccb30fbb0db

Susman-Stillman, A., Bailey, A. E., & Webb, C. (2014). *The state of early childhood assessment: Practices and professional development in Minnesota*. Minneapolis, MN: University of Minnesota.

Taylor, K. (2013). WaKIDS assessment data: State overview *WERA Educational Journal, 5*(2), 4–8.

Tindal, G., Irvin, P. S., Nese, J. F. T., & Slater, S. (2015). Skills for children entering kindergarten. *Educational Assessment, 20*, 197–319.

Torgeson, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator, Spring/Summer*, 1–8.

U.S. Department of Education. (2011). *Race to the Top—Early Learning Challenge application for initial funding* (CFDA No. 84.412). Washington, DC: Author.

U.S. Department of Education. (2013). *Frequently asked questions for the EAG-KEA 2013 competition for FY 2012 funds*. Washington, DC: Author.

U.S. Department of Education & U.S. Department of Health and Human Services. (2014). *The Race to the Top—Early Learning Challenge year two progress report*. Washington, DC: Author.

Wakabayashi, T., & Beal, J. A. (2015). Assessing school readiness in children: Historical analysis, current trends. In O. N. Saracho (Ed.), *Contemporary perspectives on research in assessment and evaluation in early childhood education* (pp. 69–91). Charlotte, NC: Information Age.

Washington Office of Superintendent of Public Instruction. (2014). *Recommendations of the Washington Kindergarten Inventory of Developing Skills (WaKIDS) workgroup*. Olympia, WA: Author.

Washington Office of Superintendent of Public Instruction. (2015). *Changes to 2015 WaKIDS whole child assessment*. Olympia, WA: Author.

Washington Office of Superintendent of Public Instruction. (2016). *Assessment and student information/teaching and learning/early learning* (Bulletin No. 012-16). Olympia, WA: Author.

Washington Office of Superintendent of Public Instruction. (2017). *WaKIDS*. Retrieved from http://www.k12.wa.us/wakids/

Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education—or whose score is it anyway? *Early Childhood Research Quarterly, 27*, 46–54.

Weisenfeld, G. G. (2016). *Using Teaching Strategies GOLD within a kindergarten entry assessment system* (CEELO FastFact). New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.

Weisenfeld, G. G. (2017a). *Assessment tools used in kindergarten entry assessments (KEAs): State scan*. New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.

Weisenfeld, G. G. (2017b). *Implementing a kindergarten entry assessment (KEA) system*. New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.

Weiss, C. H. (1979, September–October). The many meanings of research utilization. *Public Administration Review,* pp. 426–431.

WestEd. (2014). *Ready for kindergarten: Kindergarten Readiness Assessment technical report*. Sausalito, CA: Author.

WestEd. (2015). *Ready for kindergarten: Kindergarten Readiness Assessment technical report addendum*. Sausalito, CA: Author.

WestEd Center for Child and Family Studies. (2017). *Summary of instrument views for the Kindergarten Individual Development Survey (KIDS)*. Sausalito, CA: Author.

WestEd Center for Child and Family Studies Evaluation Team. (2013). *The DRDP-SR (2012) instrument and its alignment with the common core state standards for kindergarten*. Sausalito, CA: Author.

WestEd Center for Child and Family Studies & University of California, Berkeley Evaluation and Assessment Research Center. (2017). *Selection criteria for the KIDS 14 state readiness measures within 3 subsets*. Sausalito, CA: Author.

WIDA. (2012). *2012 amplification of the English language development standards: Kindergarten–Grade 12*. Madison, WI: Board of Regents of the University of Wisconsin System.

Wiggins, O. (2014, October 13). Evaluating Md. kindergartners has become a one-on-one mission. *Washington Post*. Retrieved from https://www.washingtonpost.com/local/education/marylands-kindergartners-face-important-early-test-of-their-readiness-for-school/2014/10/31/62a11caa-60fd-11e4-8b9e-2ccdac31a031_story.html?tid=a_inl&utm_term=.8a075a8b09fb

Wright, R. J. (2010). *Multifaceted assessment for early childhood education*. Thousand Oaks, CA: Sage.

Wylie, E. C. (2017). *Winsight assessment system: Preliminary theory of action* (Research Report No. RR-17-26). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12155

Yazejian, N., & Bryant, D. (2013). Embedded, collaborative, comprehensive: One model of data utilization. *Early Education and Development, 24*, 68–70.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.

Zellman, G. L., & Karoly, L. A. (2012). *Moving to outcomes: Approaches to incorporating child assessments into state early childhood quality rating and improvement systems*. Santa Monica, CA: RAND.

Zweig, J., Irwin, C. W., Kook, J. F., & Cox, J. (2015). *Data collection and use in early childhood education programs: Evidence from the northeast region* (Report No. REL 2015-084). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands.

Appendix: Data Sources

Delaware Early Learner Survey

Delaware Department of Education. (2016a). *2016–17 Delaware Early Learner Survey implementation window dates: School district schedule*. Retrieved from http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/432/2016-17%20DE-ELS%20District%20Implementation%20Window%20Dates.pdf

Delaware Department of Education. (2016b). *DE-ELS frequently asked questions 2016–2017*. Dover, DE: Author.

Delaware Department of Education. (2016c). *DE-ELS updates and improvements 2016–2017*. Dover, DE: Author.

Delaware Department of Education. (2016d). *Delaware Early Learner Survey 2016–2017 fact sheet*. Dover, DE: Author.

Delaware Department of Education. (2016e). *Guidance for surveying dual- and English-language learners*. Dover, DE: Author.

Delaware Department of Education. (2016f). *Guidelines for understanding least restrictive environments*. Dover, DE: Author.

Delaware Department of Education. (2017). *Delaware Early Learner Survey updates 2017–2018 school year*. Dover, DE: Author.

Delaware Office of Early Learning. (2016). *Delaware Early Learner Survey recorded orientation presentation*. Dover, DE: Delaware Department of Education.

Early Learning Challenge Technical Assistance. (2014). *Engaging and supporting educators in the development and implementation of KEA*. Retrieved from https://elc.grads360.org/services/PDCService.svc/GetPDCDocumentFile?fileId=5738

Hanover Research. (2017). *Delaware Early Learner Survey key findings*. Arlington, VA: Author.

Office of the Governor, State of Delaware. (2016). *Race to the Top—Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Illinois Kindergarten Individual Development Survey (KIDS)

California Department of Education Child Development Division. (2012). *Desired Results Developmental Profile: School Readiness (DRDP-SR)*. Sacramento, CA: Author.

California Department of Education Child Development Division. (2015). *Desired Results Developmental Profile: Kindergarten (DRDP-K)*. Sacramento, CA: Author.

Early Learning Challenge Technical Assistance. (2015a). *National working meeting on early learning assessment: Meeting summary*. Retrieved from https://elc.grads360.org/#program/national-working-meeting-on-early-learning-assessment

Illinois State Board of Education. (2015). *Kindergarten Individual Development Survey*. Springfield, IL: Author.

Illinois State Board of Education. (2017). *An overview on measures and data reporting*. Springfield, IL: Author.

Illinois State Board of Education. (n.d.). *KIDS Kindergarten Individual Development Survey frequently asked questions (FAQ)*. Retrieved from https://www.illinoiskids.org/content/faq-0

Illinois State Board of Education Division of Early Childhood Education. (n.d.). *Illinois early learning standards: Kindergarten*. Springfield, IL: Author.

Kriener-Althen, K., Burmester, K., & Mangione, P. (2015). *Bridging early learning, kindergarten, and first-grade content standards through assessment*. Paper presented at the annual meeting of the American Education Research Association, Chicago.

Kriener-Althen, K., & Hernandez, R. (2015). *KIDS school readiness brief: KIDS administration schedule informed by teacher input*. Paper presented at the annual meeting of the American Education Research Association, Chicago.

Office of Governor Pat Quinn. (2011). *State of Illinois Race to the Top—Early Learning Challenge application for initial funding: From birth to kindergarten and beyond*. Springfield, IL: Author.

Office of the Governor, State of California. (2016). *Race to the Top—Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Illinois. (2016). *Race to the Top—Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Illinois. (2017). *Race to the Top—Early Learning Challenge 2016 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

WestEd Center for Child and Family Studies. (2017). *Summary of instrument views for the Kindergarten Individual Development Survey (KIDS)*. Sausalito, CA: Author.

WestEd Center for Child and Family Studies and University of California, Berkeley Evaluation and Assessment Research Center. (2017). *Selection criteria for the KIDS 14 state readiness measures within 3 subsets*. Sausalito, CA: Author.

WestEd Center for Child and Family Studies Evaluation Team. (2013). *The DRDP-SR (2012) instrument and its alignment with the Common Core State Standards for kindergarten*. Sausalito, CA: Author.

Maryland/Ohio Kindergarten Readiness Assessment

Johns Hopkins School of Education Center for Technology in Education & WestEd. (2015a). *Maryland, ready for kindergarten assessments: Guidelines on allowable supports for the Kindergarten Readiness Assessment*. Columbia, MD: Authors.

Johns Hopkins School of Education Center for Technology in Education & WestEd. (2015b). *Ohio ready for kindergarten assessments: Guidelines on allowable supports for the Kindergarten Readiness Assessment*. Columbia, MD: Authors.

Maryland State Department of Education. (2015a). *Readiness matters! The 2014–2015 Kindergarten Readiness Assessment report*. Baltimore, MD: Author.

Maryland State Department of Education. (2015b). *Report on the Kindergarten Readiness Assessment (KRA): Joint chairmen's report (HB70)*. Baltimore, MD: Author.

Maryland State Department of Education. (2016a). *Readiness matters: Informing the future*. Baltimore, MD: Author.

Maryland State Department of Education. (2016b). *Readiness matters! The 2015–2016 Kindergarten Readiness Assessment report*. Baltimore, MD: Author.

Maryland State Department of Education. (2017a). *Readiness matters: Informing the future*. Baltimore, MD: Author.

Maryland State Department of Education. (2017b). *Student testing calendar school year 2017–2018*. Baltimore, MD: Author.

Maryland State Department of Education. (2017c). *The 2016–2017 Kindergarten Readiness Assessment technical report*. Baltimore, MD: Author.

Maryland State Department of Education Division of Early Childhood Development. (2015). *Kindergarten Readiness Assessment administration guide version 1.5*. Baltimore, MD: Author.

Maryland State Education Association. (2014). *MSEA report and recommendations on the Kindergarten Readiness Assessment*. Annapolis, MD: Author.

Maryland State Education Association. (2016). *Maryland General Assembly changes kindergarten readiness assessment to a sampling test*. Retrieved from http://www.marylandeducators.org/press/maryland-general-assembly-changes-kindergarten-readiness-assessment-sampling-test

Office of Early Learning and School Readiness, Ohio Department of Education. (2016). *Ohio annual report on the Kindergarten Readiness Assessment*. Columbus, OH: Author.

Office of the Governor, State of Maryland. (2014). *Race to the Top—Early Learning Challenge 2013 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Maryland. (2015). *Race to the Top—Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Maryland. (2016). *Race to the Top—Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Ohio. (2014). *Race to the Top Early Learning Challenge 2013 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Ohio. (2015). *Race to the Top Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Ohio. (2016). *Race to the Top Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Ohio Department of Education. (2015). *Update on Kindergarten Readiness Assessment reductions*. Columbus, OH: Author.

Ohio Department of Education. (2016). *KRA resources for teachers: 8/15/2016*. Retrieved from http://education.ohio.gov/Topics/Early-Learning/Early-Learning-Program-Updates/August-2016-1/KRA-Resources-for-Teachers

Ohio Department of Education. (2017). *Ohio's Kindergarten Readiness Assessment (KRA) fact sheet*. Columbus, OH: Author.

Schachter, R. E., Strang, T. M., & Piasta, S. B. (2015). *Using the new Kindergarten Readiness Assessment: What do teachers and principals think?* Columbus, OH: Crane Center for Early Childhood Research and Policy, Ohio State University.

Schachter, R. E., Strang, T. M., & Piasta, S. B. (2017). Teachers' experiences with a state-mandated kindergarten readiness assessment. *Early Years*. https://doi.org/10.1080/09575146.2017.1297777

Weisenfeld, G. G. (2017b). *Implementing a kindergarten entry assessment (KEA) system*. New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.

WestEd. (2014). *Ready for kindergarten: Kindergarten Readiness Assessment technical report*. Sausalito, CA: Author.

WestEd. (2015). *Ready for kindergarten: Kindergarten Readiness Assessment technical report addendum*. Sausalito, CA: Author.

Wiggins, O. (2014, October 13). Evaluating Md. kindergartners has become a one-on-one mission. *Washington Post*.

North Carolina Kindergarten Entry Assessment

Ferrara, A. M., & Lambert, R. G. (2015). *Findings from the 2014 North Carolina kindergarten entry formative assessment pilot*. Charlotte, NC: Center for Educational Measurement and Evaluation, University of North Carolina at Charlotte.

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, NC: Center for Educational Measurement and Evaluation, University of North Carolina at Charlotte.

Ferrara, A. M., Merrill, E., Lambert, R. G., & Baddouh, P. G. (2017). *Effects of district specific professional development on teacher perceptions of a statewide kindergarten formative assessment*. Paper presented at the 2017 annual meeting of the American Educational Research Association, San Antonio, TX.

North Carolina K–3 Assessment Think Tank. (2013). *Assessment for learning and development in K–3*. Raleigh, NC: Author.

North Carolina Office of Early Learning. (2015a). *K–3 formative assessment: Kindergarten. District implementation teams: Facilitator's guide. Interpreting and using evidence to guide instruction. Engaging families in the K–3 formative assessment process*. Raleigh, NC: Author.

North Carolina Office of Early Learning. (2015b). *K–3 formative assessment: Kindergarten. District implementation teams: Facilitator's guide. Using the formative assessment process with the technology platform*. Raleigh, NC: Author.

North Carolina Office of Early Learning. (2015c). *NC construct progressions and situations*. Raleigh, NC: Author.

North Carolina Office of Early Learning. (2017a). *Construct progression comparison document*. Raleigh, NC: Author.

North Carolina Office of Early Learning. (2017b). *NC construct progressions and situations*. Raleigh, NC: Author.

Office of the Governor, State of North Carolina. (2013). *Race to the Top—Early Learning Challenge 2012 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of North Carolina. (2014). *Race to the Top—Early Learning Challenge 2013 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of North Carolina. (2015). *Race to the Top—Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of North Carolina. (2016). *Race to the Top—Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Pruette, J. (2016, April 5). *KEA implementation, Phase II* (memorandum). Raleigh, NC: Office of Early Learning.

Pruette, J. (2017, March 16). *NC kindergarten entry assessment* (memorandum). Raleigh, NC: Office of Early Learning.

Public Schools of North Carolina. (2015). *District implementation teams: Facilitator's guide; using the formative assessment process with the technology platform*. Raleigh, NC: Author.

Public Schools of North Carolina. (2017). *NC K–3 formative assessment process: KEA updates*. Retrieved from http://rtt-elc-k3assessment.ncdpi.wikispaces.net/file/view/1.%20CCSA%20KEA%20Updates%202017_FOR%20POSTING%20ON%20WIKI.pdf/610596145/1.%20CCSA%20KEA%20Updates%202017_FOR%20POSTING%20ON%20WIKI.pdf

Oregon Kindergarten Assessment (easyCBM Literacy and Math Subtests; Child Behavior Rating Scale)

Early Learning Division, Oregon Department of Education. (2014). *Race to the Top Early Learning Challenge 2013 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Early Learning Division, Oregon Department of Education. (2015). *Race to the Top Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Early Learning Division, Oregon Department of Education. (2016). *Race to the Top Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Early Learning Division, Oregon Department of Education. (2017). *Race to the Top Early Learning Challenge 2016 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Furrer, C., & Green, B. L. (2013). *Oregon Kindergarten Assessment fall 2012 pilot process evaluation*. Portland, OR: Center for Improvement of Child and Family Services, Portland State University.

Irvin, P. S., Tindal, G., & Slater, S. (2017). *Examining the factor structure and measurement invariance of a large-scale kindergarten entry assessment*. Paper presented at the annual conference of the American Educational Research Association, San Antonio, TX.

Office of Assessment and Accountability, Oregon Department of Education. (2016a). *Oregon kindergarten assessment specification 2016–2017*. Salem, OR: Author.

Office of Assessment and Accountability, Oregon Department of Education. (2016b). *Preliminary test administration manual 2016–2017 school year*. Salem, OR: Author.

Office of Teaching, Learning, and Assessment, Oregon Department of Education. (2017). Oregon kindergarten assessment specifications 2017–18. Salem, OR: Author.

Oregon Department of Education. (2014). *Guidance for locating a Spanish bilingual assessor for the Oregon Kindergarten Assessment*. Retrieved from http://www.ode.state.or.us/news/announcements/announcement.aspx?=9909.

Oregon Department of Education. (2015a). *Decision matrix for the early Spanish literacy assessment*. Salem, OR: Author.

Oregon Department of Education. (2015b). *Oregon Kindergarten Assessment report overview*. Salem, OR: Author.

Oregon Department of Education. (2017a). *2017–18 kindergarten assessment early literacy and early math training: Required for DTCs, STCs, and TAs* [PowerPoint presentation]. Salem, OR: Author.

Oregon Department of Education. (2017b). *Test administration manual: 2017–18 school year*. Salem, OR: Author.

Tindal, G., Irvin, P. S., Nese, J. F. T., & Slater, S. (2015). Skills for children entering kindergarten. *Educational Assessment, 20*, 197–319.

Pennsylvania Kindergarten Entry Inventory

Commonwealth of Pennsylvania Governor's Office. (2015). *Race to the Top Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Commonwealth of Pennsylvania Governor's Office. (2016). *Race to the Top Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Howard, E. C., Dahlke, K., Tucker, N., Liu, F., Weinberg, E., Williams R.,. .. & Brumley, B. (2017). *Evidence-based Kindergarten Entry Inventory for the commonwealth: A journey of ongoing improvement*. Washington, DC: American Institutes for Research.

Pennsylvania Department of Education. (2015). *Pennsylvania Kindergarten Entry Inventory cohort 1 (2014) summary report*. Harrisburg, PA: Author.

Pennsylvania Department of Education. (2016). *Pennsylvania Kindergarten Entry Inventory cohort 2 (2015) summary report*. Harrisburg, PA: Author.

Pennsylvania Office of Child Development and Early Learning. (2013). *The 2011 SELMA pilot*. Harrisburg, PA: Commonwealth of Pennsylvania, Department of Education, Department of Public Welfare.

Pennsylvania Office of Child Development and Early Learning. (2014a). *The 2012 Pennsylvania Kindergarten Entry Inventory pilot*. Harrisburg, PA: Commonwealth of Pennsylvania, Department of Education, Department of Public Welfare.

Pennsylvania Office of Child Development and Early Learning. (2014b). *The 2013 Pennsylvania Kindergarten Entry Inventory pilot*. Harrisburg, PA: Commonwealth of Pennsylvania, Department of Education, Department of Public Welfare.

Pennsylvania Office of Child Development and Early Learning. (2015). *A guide to using the Pennsylvania Kindergarten Entry Inventory*. Harrisburg, PA: Author.

Pennsylvania Office of Child Development and Early Learning. (2017). *A guide to using the Pennsylvania Kindergarten Entry Inventory*. Harrisburg, PA: Author.

Washington Kindergarten Inventory of Developing Skills (WaKIDS)

Build Initiative. (2012). *Kindergarten entry assessment (KEA) question and answer session: WaKIDS: Washington State's KEA process*. Retrieved from http://www.buildinitiative.org/WhatsNew/ViewArticle/tabid/96/ArticleId/676/Kindergarten-Entry-Assessment-KEA-Question-and-Answer-Session-WaKIDS-Washington-States-KEA-Process.aspx

Butts, R. (2013). *Recommendations of the Washington Kindergarten Inventory of Developing Skills (WaKIDS) workgroup*. Olympia, WA: Office of Superintendent of Public Instruction Early Learning Office.

Butts, R. (2014). *Recommendations of the Washington Kindergarten Inventory of Developing Skills (WaKIDS) workgroup*. Olympia, WA: Office of Superintendent of Public Instruction Early Learning Office.

Dorn, R. I. (2013, August 22). *Assessment and student information/early learning education* (Memorandum no. 37-13 M). Olympia, WA: Office of the Superintendent of Public Instruction.

Dorn, R. I. (2016, April 11). *Assessment and student information/teaching and learning/early learning* (Bulletin no. 012 – 16). Olympia, WA: Office of the Superintendent of Public Instruction.

Joseph, G. E., Cevasco, M., Lee, T., & Stull, S. (2010). *WaKIDS pilot: Preliminary report*. Seattle, WA: Childcare Quality & Early Learning Center for Research and Training, University of Washington.

Office of the Governor, State of Washington. (2013). *Race to the Top Early Learning Challenge 2012 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Washington. (2014). *Race to the Top Early Learning Challenge 2013 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Washington. (2015). *Race to the Top Early Learning Challenge 2014 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Office of the Governor, State of Washington. (2016). *Race to the Top Early Learning Challenge 2015 annual performance report*. Washington, DC: U.S. Department of Health and Human Services and Department of Education.

Schilder, D. (2015). *Washington State Race to the Top—Early Learning Challenge evaluation report*. Boston, MA: Build Initiative.

Soderberg, J. S., Stull, S., Cummings, K., Nolen, E., McCutchen, D., & Joseph, G. (2013). *Inter-rater reliability and concurrent validity study of the Washington Kindergarten Inventory of Developing Skills (WaKIDS)*. Seattle, WA: University of Washington.

Taylor, K. (2013). WaKIDS assessment data: State overview. *WERA Educational Journal, 5*(2), 4 – 8.

Washington Office of Superintendent of Public Instruction. (2014). *Recommendations of the Washington Kindergarten Inventory of Developing Skills (WaKIDS) workgroup*. Olympia, WA: Author.

Washington Office of Superintendent of Public Instruction. (2015). *Changes to 2015 WaKIDS whole child assessment*. Olympia, WA: Author.

Washington Office of Superintendent of Public Instruction. (2016). *Assessment and student information/teaching and learning/early learning* (Bulletin no. 012–16). Olympia, WA: Author.

Washington Office of Superintendent of Public Instruction. (2017). *WaKIDS*. Retrieved from http://www.k12.wa.us/wakids/

Weisenfeld, G. G. (2017b). *Implementing a kindergarten entry assessment (KEA) system*. New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.

Suggested citation:

Ackerman, D. J. (2018). *Real world compromises: Policy and practice impacts of kindergarten entry assessment-related validity and reliability challenges* (Research Report No. RR-18-13). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12201

Action Editor: James Carlson

Reviewers: Brent Bridgeman and Don Powers