



Assessment and Technology: Mapping Future Directions in the Early Childhood Classroom

Michelle M. Neumann^{1*}, Jason L. Anthony^{2*}, Noé A. Erazo² and David L. Neumann³

¹ School of Education and Professional Studies, Griffith University, Southport, QLD, Australia, ² Rightpath Research and Innovation Center, Child & Family Studies, University of South Florida, Tampa, FL, United States, ³ School of Applied Psychology, Griffith University, Southport, QLD, Australia

OPEN ACCESS

Edited by:

Chad M. Gotch,
Washington State University,
United States

Reviewed by:

Timothy Mark O'Leary,
University of Melbourne, Australia
Mary Roduta Roberts,
University of Alberta, Canada

*Correspondence:

Michelle M. Neumann
m.neumann@griffith.edu.au
Jason L. Anthony
jasonanthony@usf.edu

Specialty section:

This article was submitted to
Assessment, Testing and Applied
Measurement,
a section of the journal
Frontiers in Education

Received: 27 April 2019

Accepted: 04 October 2019

Published: 18 October 2019

Citation:

Neumann MM, Anthony JL, Erazo NA
and Neumann DL (2019) Assessment
and Technology: Mapping Future
Directions in the Early Childhood
Classroom. *Front. Educ.* 4:116.
doi: 10.3389/feduc.2019.00116

The framework and tools used for classroom assessment can have significant impacts on teacher practices and student achievement. Getting assessment right is an important component in creating positive learning experiences and academic success. Recent government reports (e.g., United States, Australia) call for the development of systems that use new technologies to make educational assessment more efficient and useful. The present review discusses factors relevant to assessment in the digital age from the perspectives of assessment for learning (AfL) and assessment of learning (AoL) in the early childhood classroom. Technology offers significant avenues to enhance test administration, test scoring, test reporting and interpretation, and link with curriculum to individualize learning. We highlight unique challenges around issues of developmental appropriateness, item development, psychometric validation, and teacher implementation in the use of future assessment systems. However, success will depend upon close collaboration between educators, students, and policy makers in the design, development, and utilization of technology-based assessments.

Keywords: assessment, technology, teacher, student, education policy, early childhood education

INTRODUCTION

Assessment plays an important role in the teaching-learning process and it is a powerful tool for enhancing student achievement and facilitating societal progress (Broadfoot and Black, 2004; Hodges et al., 2014). In this twenty-first century, innovative technologies have the potential to deliver better quality educational assessments that are more useful for teachers and that more readily benefit student learning (Koomen and Zoanetti, 2018). This view is echoed by Gonski (2018) who urges educators to “use new technology not for its own sake, but to adopt ways of working that are more efficient and effective” (p. 99). Beyond commonplace technologically-supported survey methodologies, numerous new technologies offer exciting opportunities for educational assessment. These include touch screens with drag and drop and multi-touch features, augmented reality (AR), virtual reality (VR), mixed reality (MR), robots, and behavioral monitoring (e.g., voice recognition, eye gaze, face recognition, touchless user interface). It is at this nexus where innovative education theory, psychology, computer science, and engineering can combine to optimize classroom assessment practices and provide clear links between assessment, teaching, and learning.

The present review examines technology in classroom assessment from the perspective of students, educators, and administrators. Classroom assessment refers to a practice wherein teachers

use assessment data from a variety of tools or products to document and enhance student learning (Randel and Clark, 2013). While commonly used tools include teacher-made tests, the current review focuses on externally produced standardized tests by national, state, and district level assessment developers as well as commercial developers. Assessment can be conceptualized in two ways: as facilitating the learning process and as summarizing the current state of knowledge in students. Technology has the potential to enhance both applications. Moreover, technology offers significant advantages across the different stages of assessment, from test administration to linking data to the curriculum. However, concerns in using technology-based assessment have also been raised around developmental appropriateness, item development, psychometric validation, and teacher training. The present review examines these issues, with a focus on technology-based assessment for education in the early years. The early childhood classroom, for the purposes of this review, includes kindergarten and the preparatory year. In some regions, early childhood may also refer to the 2 years prior to and the year following kindergarten. Following an overview of assessment processes in education, we examine the use of technology in assessment before concluding with future areas in need of development.

UNDERSTANDING ASSESSMENT PROCESSES IN EDUCATION

In the educational context, assessment is broadly conceptualized as an ongoing process of gathering evidence of learning, interpreting it, and acting on this evidence to improve future learning and performance (Stiggins, 2002; Bennett, 2011). In this respect, assessment is understood as a social-cultural practice or activity (Broadfoot and Black, 2004; Looney et al., 2018; Silseth and Gilje, 2019). It is embedded in the teaching and learning process which is mediated by the tools used in assessment. Furthermore, the processes used in assessment are closely linked with the social interaction of learners and teachers, with the construction of knowledge achieved by a novice-expert relationship. The quality and individualized feedback to students is also integral to the process (Sadler, 1989; Heritage, 2007). As such, assessment that incorporates both social and individualized perspectives is likely to help student learning (Hodges et al., 2014). Successful assessment systems of the future will closely embody the needs and perspectives of teachers and their students.

The application of assessment within this broader framework generally falls within three categories, namely diagnostic assessment, summative assessment, and formative assessment. These three types are distinguished by their purposes, timing, to whom they are administered, and in test construction and design. However, there can be instances when the same test is used for more than one application, which may not necessarily be appropriate if the test was not designed for this. Diagnostic assessments are designed to thoroughly assess achievement in a given domain and all relevant subdomains. Diagnostic reading tests, for example, assess children's phonological awareness, graphophonemic knowledge, reading fluency, and reading

comprehension. Diagnostic tests are administered to individuals who are struggling to learn or who have been deemed at-risk of academic failure. Results from well-designed diagnostic tests help inform educators and special educators what to teach and how to teach. Because diagnostic tests are usually designed to classify students and to determine access to special services, they are rigorously developed and administered in ways that assure that the test scores and their interpretation have high degrees of reliability, validity, and fairness. As such, they are lengthy and often require some expertise on the part of the assessor.

Summative assessments are designed to quantify how much one has achieved to date in a given academic domain and their purpose is *assessment of learning* (AoL). Summative assessments are standardized tests that are usually administered to all students in a given grade, school, school district, state, or country. AoL occurs at a specific point in time where achievement to date is to be quantified. This is typically at the end of an academic school year, completion of a course, or immediately following an intervention program. Examples include final exams, school district administered standardized tests, the Graduate Record Examinations (GRE), the National Assessment of Educational Progress (NAEP), and the National Assessment Program Literacy and Numeracy (NAPLAN) [Australian Curriculum, Assessment and Reporting Authority (ACARA), 2013]. Results from summative assessments may be shared with students, parents, teachers, administrators, and evaluators. These consumers use the indices of overall student achievement to make evaluative judgements against predetermined standards. In recent years, AoL is being increasingly used for high stakes accountability purposes (Stiggins, 2002; Heritage, 2007). For example, in much of the United States, AoL data are used to rank order public schools, determine teacher and principal salaries, decide whether to retain or terminate principals and school district administrators, determine the need for third party takeover of public schools, and defund publicly funded early childhood education programs (Darling-Hammond, 2004; Neal, 2011).

Formative assessments are designed to efficiently measure how well students are responding to instruction in a specific subdomain of achievement and to indicate if instructional modifications are warranted. Their purpose is *assessment for learning* (AFL). AFL does not aim to quantify overall achievement. Instead, its purpose is to generate data useful for guiding instruction. That is, AFL has a focus on the integration of assessment activities into the teaching and learning process. In AFL, test results provide immediate feedback to teachers and students about how much of the recently taught material has been learned. The results are used by teachers to inform lesson planning (Sadler, 1989). Wiliam (2011) notes that educators who use formative assessment must have a strong understanding of what the learner knows, where the learner is going, and how to get there. The feedback afforded to educators and students through AFL serves to guide the learner through individualized teaching approaches that optimize student learning (Wiliam, 2011). It helps students improve as they work to attain higher levels of performance to create new knowledge and highlights the important relationship between classroom assessment practice, learning, and use of assessment evidence to guide instruction.

Early identification, targeted instruction, monitoring of children's learning, and data-driven instructional changes are key components of programs that close achievement gaps. AfL takes many forms and can inform each of these components. For example, for educators to provide targeted instruction, a student's mastery of taught skills and their (sub)domain specific learning must be regularly assessed to determine progress toward desired outcomes. Skills mastery tests, traditionally called curriculum-based measures, are one form of AfL and tests assess the extent to which a child has learned specific skills taught in a given curriculum. Skills mastery tests are brief, closely linked to the curriculum, and administered frequently (e.g., weekly spelling tests). Students' performance on skills mastery tests helps educators appropriately pace their progress through a given curriculum. These tests are necessary but not sufficient for guiding instruction because mastery of a particular skill does not necessarily lead to mastery of that academic domain or subdomain (Fuchs, 2004; Shapiro et al., 2004; VanDerHeyden, 2005). For example, a student who can read "-at" word families may still have difficulty reading a passage that incorporates a variety of rhymes and word structures.

A useful AfL approach includes both mastery tests and General Outcome Measures (GOM; Deno, 1985, 1997). GOMs are broader in item content than mastery tests, and they are not usually linked to a specific curriculum. GOMs are usually administered to all students in a classroom, grade, or school district at predefined increments of time. For example, universal benchmarking often occurs three or four times per school year. GOMs are also administered more frequently to those students who are receiving more frequent or more intensive intervention. The potential strengths of GOMs include brevity and ease of administration, alternate forms that allow frequent re-administration, sensitivity to learning, and implications for grouping children and modifying instruction. These assets make GOMs a fitting approach for monitoring students' progress and evaluating their responsiveness to instruction. GOMs help teachers evaluate students' level and rate of achievement, determine needs for instructional change, set appropriate short- and long-term goals, and monitor progress relative to peers or criterion-based benchmarks (Shapiro et al., 2004; VanDerHeyden, 2005; Busch and Reschly, 2007). Thus, GOMs have come to the forefront of educational assessment with the emergence of response to intervention (RTI) frameworks for service provision and identification of children with learning difficulties.

RTI is a framework for linking AfL to instruction through data-based problem-solving. RTI includes an effective core curriculum, increasingly intense tiers of instruction for underperforming students, integrated assessment including universal screening, benchmarking, mastery tests, and progress monitoring, and use of assessment results to guide instruction. RTI can be implemented by teachers and when done so improves student outcomes (Fuchs et al., 1984, 1989; Graney and Shinn, 2005; Heritage, 2007; VanDerHeyden et al., 2007) and is satisfying for teachers (Hayward and Hedge, 2005). A meta-analysis reported impressive mean effect sizes of 1.02 for field-based studies and 1.54 for

university-based studies evaluating RTI implementations (Burns et al., 2005). Practice guides from the U.S. Department of Education's Institute of Education Science (IES) conclude that there is strong evidence for effectiveness of RTI (Gersten et al., 2008, 2009).

Application of New Technologies for Assessment

The application of technology may provide one avenue for resolving the intricacies of classroom assessment in the twenty-first century. Research between assessment and classroom learning help to refine technology-based supports and theoretical models of assessment, teaching, and learning processes (Black and Wiliam, 1998; Heritage, 2018). To develop the next generation of technology-based assessments, test developers will need to consider the perspectives of policy makers interested in content standards, teachers interested in AoL and AfL, and assessment experts interested in the results collected (National Research Council, 2010, p. 21). The use of technology in classroom assessment promises advanced features not possible with paper-and-pencil tests, such as faster student feedback and computer-generated next steps that allow teachers to make real-time data-driven decisions to inform their instructional changes. In order to realize such insightful and sophisticated technology, attention to student-centered assessment and instructionally tractable assessments is highly recommended (Russell, 2010; Wiliam, 2010). A collaborative approach to test development will improve the implementation process for using computer-based assessments in the classroom.

There are other ways that test developers can advance knowledge in areas such as early childhood classroom assessment, such as designing assessments that align with the five dimensions of innovation for computerized tests (Parshall et al., 2000). The field is ripe for exploration in the area of design features for children, such as item formats, response action, media inclusion, interactivity, and use of scoring algorithms. Research on computer use in young children is still in its infancy and empirical research is newly emerging (Clements and Sarama, 2003; Labbo and Reinking, 2003; Chen and Chang, 2006; Schmid et al., 2008). Technology can be used to enhance children's learning experience in the classroom, which is also expected to prepare active and informed citizens in a competitive global economy [Ministerial Council on Education, Employment Training, and Youth Affairs (MCEETYA), 2008]. The development of innovative computer-based assessments for children will require a rich understanding of developmentally appropriate design features, content expertise, implementation science, measurement, and an understanding for what students and teachers need.

Developmental Appropriateness

The digital age has initiated a generational shift where children are increasingly likely to have openhanded access to technology. Approximately two-thirds of USA citizens now own a smartphone (Pew Research Center, 2015) and ongoing research suggests that even some children from low-income, minority communities have near universal access to mobile devices

(Ojanen et al., 2015). The American Academy of Pediatrics (AAP) currently recommends that children younger than 18 months should avoid screen media and children ages 2 to 5 should limit their screen time to 1 h per day of quality programs (American Academy of Pediatrics, 2016). While the research evidence on child and technology use continues to grow, studies on children's computer interventions have demonstrated promise in areas like language and literacy (Lankshear and Knobel, 2003; Burnett, 2010; Neumann, 2018; Neumann et al., 2019). Arguably the biggest factor relating to developmental appropriateness is the nature of the technology itself.

Research has repeatedly shown that young children can experience difficulty manipulating a computer mouse when performing drag and drop sequences due to their limited motor skills, eye-hand coordination, and the size of their hand relative to the mouse (Joiner et al., 1998; Hourcade et al., 2004; Donker and Reitsma, 2007). Instead, the use of touch screen tablets in education and assessment for young children is recommended. Touch screen tablets can be used by young children and children with special needs who may lack the fine motor skills to effectively use a standard keyboard or mouse (Neumann and Neumann, 2018). Using multimodal features, touch screen devices offer opportunities to administer tests in ways that can facilitate the assessment process (Lu et al., 2017).

With the widespread use of touch screen devices, feasibility research on the developmental appropriateness of children's tablet use is underway. Early findings suggest that 2-year-olds can perform tap and drag gestures when using touch screen devices, and 3-year-olds can tap, drag, free rotate, as well as drag and drop (Aziz et al., 2014). Touch screen tablets offer different ways for students to interact with the screen and thus allow for test items to conform to many different item types. Children can use their fingers to draw, tap to highlight objects, swipe objects away, tap and drag objects to other places on the screen, pinch to zoom in and out, twist to rotate objects, and scroll up and down a screen. This physical interaction can also create a testing situation that is more engaging for children than traditional paper-and-pencil tests (Woloshyn et al., 2017).

As children develop their fine motor skills and advance to writing, there is also the capability to assess handwriting using a stylus pen. A stylus pen allows children to create shapes and letters and form lines of different thickness when pressure is applied on a digital surface. Research shows that children can easily manipulate the stylus for drawing and writing and are engaged by the activity (Chang et al., 2005; Matthews and Seow, 2007). Falk et al. (2011) demonstrated the feasibility of measuring children's handwriting by using a Wacom Intuous 3 digital tablet and a custom-built pen. These digital tools measured spatial, temporal, and grip force parameters. In their sample of first and second graders, static grip was associated with lower legibility. These input methods offer a variety of ways to appropriately assess multitouch gestures and handwriting skills in older students.

Assessing toddlers using touch screen technology is the new frontier. Twomey et al. (2018) found that children as young as 2-years-old can complete a cognitive assessment using a touch screen device. A range of touch screen technologies are already

being developed and applied in classroom assessment as the preferred response action. For example, a tablet is used in the Profile of Phonological Awareness (PRO-PA) in which it provides an interface for the teacher to ask questions and enter in student responses (Carson, 2017). A tablet is also used in the validated Emergent Literacy Assessment app (ELAA) which plays pre-recorded audio to ask questions and uses a touch screen interface to collect responses from the child (Neumann and Neumann, 2018; Neumann et al., 2019). Future research is needed to enhance developmentally appropriate features of tablets to improve digital assessment experiences for young children.

Item Development

Technology-based assessments offer more variety in stimulus presentation than is available with paper-based test booklets or flip books. Touch screen tablets, computers, and virtual modalities have multimodal features to give students opportunities to strengthen learning, motivation, collaboration, engagement, and productivity and can be used for multiple formats in assessment (Woloshyn et al., 2017). The use of technology promises improved measurement of higher-order understanding and performance because of its flexibility in integrating media and exploring new item types. The criticism against current state-level assessments are that they rely heavily on multiple choice items thereby suggesting a lack of rigor (National Research Council, 2010). There is a vocal disenchantment with multiple choice items due to a reported overreliance on measuring factual knowledge rather than higher-level skills (Pellegrino and Quellmalz, 2010). A proclivity for multiple choice items in assessments has an overarching effect in the classroom as well and research suggests that teachers are more likely to rely on multiple choice items in their classrooms when year-end assessments do too (Abrams et al., 2003). Nevertheless, multiple choice test items are more efficient than open-ended items (Jodoin, 2003), easier and cheaper to develop (Stecher and Klein, 1997), equitable for children of different backgrounds (Bruder, 1993), and can be refined to measure higher-level skills (Parshall et al., 2000). For emergent readers, multiple choice items can be designed as multiple-choice graphics. The use of technology-enhanced test items in early childhood assessments is still largely untapped and a delicate balance between innovation, cost, and efficiency is needed when designing items.

To increase the rigor of multiple-choice items, test developers and teachers can use multiple choice variants. By using multiple response items, children must choose more than one answer choice to get the item right. By using ordered response items, children must choose the correct sequence for an event. Technology can enhance and facilitate the administration of these and other types of items with the use of touch screen technology as they can be used in multiple formats. For example, students can touch a hot spot on a graphic as their answer choice (O'Neill and Folk, 1996; Parshall et al., 1996, 2008; Scalise and Gifford, 2006; Becker et al., 2011). Students can also highlight texts for assessment purposes (Davey et al., 1997) and be assessed on their drawing and mark making abilities (Scalise and Gifford, 2006; Kopriva and Bauman, 2008; Boyle

and Hutchison, 2009; Dolan et al., 2010). Drag and drop features can be used to select and move objects, order objects, connect objects, and sort objects. The limits of traditional item types can be further explored with the use of touch screen technology, and in addition can be enhanced by integrating media-based features (e.g., sounds, animations).

Adapting a multiple-choice paper-based test into a computerized format is a natural evolution when transitioning to technology-based assessments, but there is a growing call for greater innovation. The use of media such as graphics, audio, and video are ideal for emergent readers who are not yet fluent readers. Recent and future advances in behavioral monitoring (e.g., eye gaze, face recognition, touchless user interface) offer exciting opportunities for even more diverse ways that students may demonstrate their learning. For example, group administered expressive tests may become a possibility to the extent that voice recognition software advances to accommodate dialectal differences and multilingual influences on articulation and tone. Similarly, gesture recognition and facial expression recognition provide additional non-verbal modalities to help reduce the reliance on verbal skills common to many traditional assessment approaches.

Incorporation of movement via animation or video clips readily supports assessment of verbs on vocabulary tests, which have always been difficult to elicit from static illustrations on traditional paper-based assessments. A study on computer-based storytelling in kindergarteners found that computer administered stories using animation, video, sounds, and music were more effective at supporting language development than computer administered stories using still images (Verhallen et al., 2006). Augmented reality (AR), virtual reality (VR), and mixed reality (MR) presentations offer highly engaging stimulus presentation in the foreground, experimental control of the background, and truly interactive means of responding. The interactive computer tasks of the future will include multiple modes of assessment and there is headway being made in the area of K-12 science assessment. Opportunities to develop interactive computer tasks should be taken when these offer advantages to static assessment modes. For example, items can be developed to indicate slow motion, scenarios that are invisible to the naked eye, hazardous situations, and the manipulation of objects (National Assessment Governing Board, 2014). Linking students within the same virtual environment through avatars may also offer the potential to assess skills requiring teamwork, cooperation, and communication.

Psychometric Validation

As new online assessment systems and related educational policies are introduced in many countries around the world (e.g., Australia, USA), it is essential that rigorous test development work, piloting of the technologies with students and educators, and testing of infrastructure are conducted prior to large scale rollout. For example, when former paper-based assessments are transitioned to digital platforms, developers must attend to comparability of the two versions in terms of content, psychometrics, construct validity and scoring. Research suggests that differences in test administration mode are more significant

than the content itself in affecting test outcomes (Bridgeman et al., 2003; Pommerich, 2004). Therefore, it is recommended that test developers conduct comparability studies of their tests at the total score level and the item level to ensure score equivalency across their paper and computerized tests. A substantial number of test items need to be produced in anticipation that extensive field testing and post-administration analysis will result in the reduction of problematic items.

Also, quality concerns should be urgently addressed in terms of the reliability and validity of the test scores (Nickerson, 1989; Koomen and Zoanetti, 2018) to ensure teacher and public confidence (Broadfoot and Black, 2004). Assessments used in education should be subject to the same rigorous validation processes as any other cognitive assessments used for psychological diagnostic purposes (e.g., intelligence tests). In this respect, there are psychometric properties of educational assessments that are particularly important.

Measurement invariance refers to the property that an assessment is stable across different groups (e.g., gender, cultural groups). Technology-based assessments should demonstrate the same invariance, which can be important given that some groups of children will have different exposure to technological devices than other groups (e.g., greater access to tablets and other devices for children from higher socioeconomic families). In a related concept, the assessment of the same construct (e.g., phonological awareness) should demonstrate invariance regardless of what technology is used or even if traditional paper-and-pencil methods are used.

Another psychometric property is test-retest reliability, which is particularly important for applications in AFL. The stability of a measure over repeated measurements is essential if teachers are to infer changes as a result of learning and plan future lessons accordingly. New technologies are needed that also allow in-depth analysis at more granular levels. Whereas, error analysis usually takes extensive time by the test administrator outside of the testing context, automated error analysis via computerized scoring is almost immediate. Focus groups with teachers have found they desire this granular level of reporting because it helps them identify learning gaps and plan accordingly (Landry et al., 2017).

Test Administration

Technology-based assessments have the potential to make the process of administration more standardized, efficient, and offer diverse ways in which children can demonstrate their knowledge. Computer administered tests with automated scoring improves ease of use and minimizes administration and scoring related errors by teachers (Foorman et al., 2008). Some of these advantages can be witnessed even when tests are administered individually, and students' answers are entered directly into a computer by a test administrator. Online tablet testing also offers several practical advantages over (online) desktop computer testing. They are compact and mobile which allows them to be utilized in a range of contexts. Children can use them on a desk or when sitting on the floor and can carry them around the classroom allowing increased flexibility for test settings and provides greater choice for individual student preferences.

However, teachers will need to provide support to young children so that they are properly engaging with the computer (Hitchcock and Noonan, 2000; Ellis and Blashki, 2004) which will improve their use of technology over time (Klein et al., 2000).

Nevertheless, the use of a consistent set of instructions and the ability to enter responses directly into the database leads to increases in efficiency and standardization. Fully computerized applications that automatically present training items, instructions, and test items, and that automatically gather student responses, optimize standardized administration. This important benefit avoids otherwise inevitable variations among test administrators like in timing and dialect, which can invalidate test scores in some types of educational assessments like tests of phonological awareness, listening comprehension, and mathematical problems. Efficiency gains are the most often cited advantage for the use of technology-based assessments in the classroom. While the automatic presentation and collection of student responses is a commonly cited example, there are other ways that technology improves efficiency.

Computer adaptive testing (CAT) is a method of administering tests that adapts to an examinee's ability (Wainer, 1990). CAT interacts with the examinee by selecting items that maximize precision of the test based on what is known about a student from his or her prior responses. Test administration is individualized as item difficulty is made easier or harder following incorrect or correct items, respectively. The tailoring of items is performed using item selection algorithms such as multidimensional adaptive testing (Luecht, 1996; Segall, 1996). Adaptive testing reduces the need to administer all the items to all children, thereby saving time. Shortening of the test may increase student engagement, thereby also increasing the degree of accuracy (Olson, 2002). CAT selects items from large item pools and tests of different lengths may be administered based on user input concerning the level of precision in scores desired in accord with the purpose of the testing. CAT is however criticized for not allowing users to review or change their answers once they have responded (Wise, 1996, 1997; Pommerich and Burden, 2000). A solution to this would be to adopt testlet-based CAT which adaptively administers a subset of items, or a "mini-test," rather than item by item to users (Wainer and Kiely, 1987; Wainer and Lewis, 1990). CAT is well-suited for efficient benchmarking and progress monitoring because subsequent administrations resume where prior administrations terminated. Examples of children's CAT include the STAR Reading assessment (Renaissance Learning, 2015) and the Smarter Balanced Assessment Consortium (2018) that monitor children's progress on a large-scale.

An efficient and intuitive assessment system is an enormous time saver for administration and provides the foundation for advanced reporting features, and is key to user satisfaction. For example, a variety of means are now available to electronically import students' names, external identifiers, grades, birthdates, sex, ethnicity, free and reduced lunch status, special education status, and English language learner status. Many state education agencies (SEA) and local education agencies (LEA) in the USA consider functionality for bulk upload of student details as a prerequisite for purchase of any new technology-based

assessment products. The most user-friendly assessment systems also allow importing of teacher, school, and district information and follow with an intuitive means for administrators at each level to specify the roles and relations among students, classes, teachers, special educators, school administrators, district administrators, and SEA administrators. These specifications are used by the technology to assign user access privileges to assure that each user only has security access to appropriate data. The demographic data and information stored and optionally edited in the user management system support the scoring and reporting functionalities, as different graphical views of data at different levels of aggregation can be provided to administrators, teachers, special educators, interventionists, and parents.

Test Scoring

In many cases, one of the major benefits of technology-based assessments is their ability to automate the collation and scoring of assessment data. This digitized or computerized scoring process enhances efficiency and accuracy. This is achieved by no longer requiring humans to perform data entry, calculate raw scores, transfer scores, search and locate the appropriate look-up tables, calculate domain scores, and perform a number of score conversions (e.g., raw score to age score, raw score to grade score, raw score to ability score, raw score to norm referenced standard score). Beyond provision of improved accuracy and efficiency of common scoring processes, technology-based assessments exclusively offer the ability to capitalize on modern psychometric models.

The scoring of most traditional educational assessments assumes that all items on a given test are equally able to index the construct of interest. However, this assumption is rarely supported by statistical analysis. For example, two parameter logistic (2PL) item response models better explain performances on tests of phonological awareness (Anthony et al., 2002, 2011), oral language (Anthony et al., 2014) and letter knowledge (Anthony, 2018) than do one parameter logistic (1PL) models. Computerized scoring can weight items by their discriminations, creating estimated ability scores with greater precision. Moreover, only computerized scoring can incorporate the most advanced psychometric models that are becoming more common in educational measurement (e.g., three parameter logistic models, graded response models). Most exciting on the horizon are psychometric item response models that consider both accuracy and latency data in estimation of student abilities, which of course, only computerized scoring could accommodate. Scoring of student assessment data that considers both accuracy and latency of student responses may in turn have different instructional implications.

Test Reporting and Interpretation

Making sense of assessment results and using them appropriately are some of the biggest challenges faced by educators and administrators whose formal education does not typically include advanced coursework in measurement. This is another area in which technology-based assessments offer significant advantages over traditional educational assessments. New technologies can support interpretation of results with tabular reports and

graphical plots of an individual's learning rate relative to a variety of pertinent reference groups. Data should be reported to educators in a way that optimizes their interpretability by considering the latest research on educators' statistical literacy, such as following established standards (Rankin, 2016). Important referent groups that help contextualize a given student's learning include the average learning rates of peers in a small group, classroom, grade, school, school district, and national norms. Student level reports may also be shared electronically with parents, if included in the user management system. Otherwise, traditional parent reports may be printed and distributed via mail or discussed during parent/teacher conferences. Reporting using digitized platforms can be enhanced through user-friendly visual graphics, graphs, and tools to track learning over time. These individualized digital reports and records will benefit teachers, parents, students, and schools. For example, if students move schools locally, nationally, or internationally their digital reports can be easily accessible and travel with them. Such digital enhancements will help inform the student's new teachers and schools of their current competencies and learning goals.

To further support educators and administrators, reporting and data visualization can occur at higher levels of aggregation, e.g., small group, classroom, grade, school, and school district. Moreover, for those systems based on very large normative samples with links to child and school level demographics, demographically adjusted reporting may also be available. This is particularly relevant for schools and school districts, for example, that serve high proportions of students with economically disadvantaged backgrounds or special needs and those that serve high proportions of dual language learners. There is also the potential that widespread electronically administered and stored assessments can create big databases to inform education policy reports and practice. However, although 95% of education leaders indicated that big data technology allows greater in-depth knowledge about student learning, many schools are slowly transitioning to cloud computing and mobile technologies (Harvard Business Review Analytical Services, 2017). Technology can assist with reporting which is essential for Afl purposes. As noted by Gonski (2018, p. 62), nationally administered standardized tests "provide a useful 'big picture' view of student learning trends across Australia and the world, but have limitations at the classroom level: they report achievement rather than growth Teachers need to have useable data about each student at their fingertips as the basic prerequisite for improving learner outcomes." These barriers limit reporting and interpretation and a greater focus on using data reporting to support individualized learning is needed.

Links to Curriculum and Individualized Learning

Technology offers much promise for supporting educators in making data-driven instructional modifications. For example, technology may help educators set realistic instructional goals that simultaneously consider a student's current proficiency level, his or her predicted growth rate, demographic characteristics of

the student, the standard error of predicted growth rate in light of test reliability and number of data points, and normative growth rates. Students' progress toward individual goals and normative benchmarks can be evaluated at each progress monitoring wave and instructional modifications made if necessary. For example, the Texas Kindergarten Entry Assessment (TX-KEA) is used by teachers as a school readiness screener in domains like listening comprehension in English and in Spanish (Anthony et al., 2017). Preschoolers listen to the prompts on headphones and answer the questions by touching a colorful illustration presented in a multiple choice format. TX-KEA also includes multiple response items that require children to touch multiple illustrated objects to get the item correct. As the field of child and technology use advances, the developmental appropriateness of technology will need to be considered beyond the early years. This is particularly the case for the more sophisticated technologies, like VR, AR, and MR technologies that are intuitive, commonplace, and authentic to the assessment process will provide more individualized educational data for teachers and students for planning learning experiences.

Some new technology-based assessments, like the Texas Kindergarten Entry Assessment, guide educators through RTI by provision of small group recommendations based on children's achievement profiles and recommendations to specific lessons in a curriculum or links to supplemental online instructional materials (Landry et al., 2017). Theoretically, computer generated instructional recommendations and links to supplemental instructional materials could be based on both error analysis and achievement profiles. Effectively taking the guesswork out of RTI is a significant provision of technology-based assessment, especially for the early childhood education workforce that is sometimes less formally schooled in linking assessment to instruction.

The recent Gonski (2018) review panel recommended the development of "a new online and on demand student learning assessment tool based on the Australian Curriculum learning progressions" (p. 66). In the context of education in the USA, Stiggins (2010, p. 763) argues that "our national assessment priority should make certain that assessments of both *of* and *for* learning are accurate in their depiction of student achievement and are used to benefit students" and recommends an online approach that includes both Afl and AoL to support teacher planning and student learning. In application, this knowledge will allow teachers to collect and share data in ways that will allow for the systematic creation of new learning experiences for students, facilitate transitions, and to evaluate the effectiveness of education policies, programs, and teaching practices. Digital links to the curriculum will enable efficient and accurate transfer of student outcome evaluations to tangible and effective learning experiences that support student's progression within their Zone of Proximal Development (Heritage, 2007).

Furthermore, technology also offers promise for supporting administrators in making data-driven changes. For example, classroom level aggregated reports may help school level administrators make decisions about allocation of limited professional development resources, limited curricular resources, and school wide supplemental services. School level and district

level aggregated reports may similarly support district level administrators and state level administrators make decisions about professional development needs, curricular needs, supplemental programming needs, and topics to address with new education policies.

Teachers and Implementation

While evidence for the benefits of AfL approaches continues to mount, research is needed around the optimal design of all forms of assessments—diagnostic, summative, and formative—and how new technologies can enhance the use of these complementary tools. Evidence centered design (ECD) is gaining ground as an assessment design and developmental framework for incorporating authentic and interactive tasks. The framework is an iterative design process that covers design, student performance, data, and test delivery with the aim of producing cost-effective tasks with clear links to the target construct (Mislevy et al., 2003). New technologies also make the design of accessible products a reality for all children, with an understanding of their abilities and preferences. The adoption of universal design can minimize the need for test accommodations by designing products that are accessible to children regardless of disabilities (Salend, 2009). Furthermore, as tablets are relatively inexpensive some schools now require that children purchase a tablet as part of a BYOD (bring your own device) program in much the same way as calculators had to be purchased. The lower price increases the potential for widespread use and applications of tablets by students and teachers. Ultimately, a student and teacher centered approach will help guide research and practice on optimal assessment design.

Teachers are being increasingly encouraged to implement new technologies in their classroom assessment practices. This pressure comes from the promise that the technologies can better meet changing stakeholder expectations, fulfill new assessment purposes, be engaging for students, deliver timely and informative results, and are flexible and efficient in administration and scoring (Bennett, 2011; Gonski, 2018; Koomen and Zoanetti, 2018). These high expectations will continue to be unmet until teachers are provided with adequate training on sound classroom assessment practices and the use of technology. Research suggests that having higher levels of assessment knowledge leads to increased use of a variety of assessment tools in the classroom (Bailey and Heritage, 2008; Popham, 2009).

Following initial applications, it was acknowledged that technology-based approaches to assessment presents challenges at the classroom and school levels (Stiggins, 2002; Heritage, 2007). Since that time, the increased use of technology has resulted in a better understanding of how to meet these challenges. Teachers must be given significant pedagogical guidance to understand new assessments and ensure school engagement and participation in use of new assessment processes (Looney et al., 2018; Van der Kleij et al., 2018). For example, teachers must be confident and comfortable applying consistent scoring procedures to collect AfL and AoL data from assessments that are clearly aligned with curriculum and instructional objectives. Communication of assessment must

also be understood clearly by students, parents, and caregivers. An important aspect of this communication is delineating the relationship between assessment and learning, and research is needed to refine sociocultural assessment theory in the context of online and mobile technologies (Baird et al., 2017).

Teachers play a key role in administering assessment and using data to inform planning for teaching and learning. As is stated in the Australian Professional Standards for Teachers (Standard 5), teachers are required to “Assess student learning, provide feedback to students on their learning, make consistent and comparable judgements, interpret student data, and report on student achievement” [Australian Institute for Teaching and School Leadership (AITSL), 2011 p. 16–17]. Teachers are also expected to develop, select, and use AfL strategies to assess student learning and provide timely effective and appropriate feedback relative to their learning goals. This approach is also reflected in position statements on assessment in USA schools. For example, in the 2001 report of the Committee on the Foundation of Assessment of the National Research Council-Recommendation 11: “The balance of mandates and resources should be shifted from an emphasis on external forms of assessment to an increased emphasis on classroom formative assessment designed to assist learning” (Stiggins, 2002, p. 763). For assessments to be effective, teachers must have a sophisticated level of knowledge of both curriculum and AfL practices (Van der Kleij et al., 2018). However, many teachers are unprepared in the use of AfL practices (Stiggins, 2002; Lopez and Pasquini, 2017) and assessment is often perceived by teachers as high stakes, rank focussed, and “something that is in competition with teaching, rather than as an integral part of teaching and learning” (Heritage, 2007, p. 140). Also, due to the increasing use of new technology in classrooms, more teacher knowledge is needed to understand the complex relationship between AfL and AoL. Without this knowledge, teachers are likely to avoid adopting new assessment practices, which may otherwise be of benefit (Stiggins, 2002).

Teachers are the front-line professionals who have the responsibility for facilitating teaching and assessment. As such, teachers would benefit from professional development activities that assist them to gain sophisticated knowledge of both the curriculum and AfL practices (Stiggins, 2002; Heritage, 2007). For example, teachers need support to plan and implement quality assessment tasks, interpret evidence, develop outcomes appropriate to assessment purpose and type, generate feedback, report, and engage students as active participants in their assessment and learning (Looney et al., 2018). AfL is often not well-understood by teachers (Deluca et al., 2012), is not strong in practice and many teachers are unprepared to make summative judgements (Lopez and Pasquini, 2017).

Furthermore, teachers can be challenged by conflicts among their belief systems, institutional structures and pressure from external accountability testing (Black and Wiliam, 1998; Dwyer, 1998). However, it has been found that teachers with stronger confidence were better at AfL in the classroom (Allinder, 1995), suggesting that enhanced self-efficacy with assessment tools and practices can be of benefit. We need to work with teachers to help them maximize the use and benefits

of assessment technologies in the classroom. However, while most teachers possess general knowledge about using new technologies in the classroom, some experience uncertainties about their capabilities for meaningfully integrating tablets, computers, and mobile devices into the classroom for teaching, assessment, and tracking student progress (Woloshyn et al., 2017). Clear evidence-based pathways are needed for the smooth transition from traditional (paper-and-pencil) to technology-based assessment so that teachers can seamlessly integrate technology into existing approaches and take advantage of technology (e.g., tablets/iPads) and its flexibility and mobility (Lu et al., 2017).

Clearly, barriers still need to be overcome in order to allow the seamless use and implementation of assessment technologies in the classroom. Time is needed to help teachers build AfL and technology skills, reflect, interpret, and develop formative assessment materials to suit their students' learning needs. Provision of professional development will assist teachers in becoming proficient and confident users of AfL practices to effectively track student progress (Lu et al., 2017), and use assessment technologies (Dwyer, 1998; Woloshyn et al., 2017). The importance of investing in pre-service training for competency in assessing student learning must continue by policy makers committed to investing in in-service professional development in a coordinated approach that provides teachers the expertise and support they need (Stiggins, 2002; Heritage, 2007).

Indeed, it has been found that successful use of education technologies (e.g., to prepare students for state assessments) in the classroom is dependent upon consistent, extensive and quality teacher professional development programs (Penuel et al., 2007; Martin et al., 2010). Long term (2 years) professional development programs that assist teachers in integrating technology for teaching and learning can change practice, support the learning of new technologies, and show how technology can help students achieve learning goals (Lawless and Pellegrino, 2007). It is important to also consider that teacher professional learning needs the emotional involvement of teachers to embrace change and new assessment innovations. The creation of collaborative support networks that shifts traditional assessment mindsets of all stakeholders is critical in raising teacher knowledge about AfL and will in turn foster student confidence in themselves as learners. Only then will a positive pathway to student growth and achievement be recognized (Wiliam, 2011).

FUTURE CONSIDERATIONS FOR TECHNOLOGY-BASED EDUCATIONAL ASSESSMENT

Digital technologies have the potential to be a powerful assessment tool for teachers and students (Woloshyn et al., 2017). Ultimately, both educators and administrators have the role to effectively integrate new technologies into the classroom. However, further work is required if the potential advantages of technology-based assessments are to be realized. It is important

that research is conducted to build an evidence base from which to establish best practices. Furthermore, there are some issues that are particularly salient for the introduction and use of technology-based assessments. These include developmental appropriateness of the technology, ensuring that test scores are valid and reliable, and ensuring that teachers are supported in the use of technology for assessments. From the perspective of educators and administrators preparing for universal testing as part of either AoL or AfL, the assessment process actually begins with documenting basic student demographic characteristics that are relevant for later scoring and interpretation of the results. This is followed by test administration, test scoring, test reporting, and test interpretation, through which the results are fed back to students and teachers. Teachers can then use this data to monitor learning and inform the design of individualized learning experiences that are linked to the curriculum. Students must be inherently involved in the process of assessment which is viewed as the continuous flow of information between the student and teacher where the process of learning and growth is of central focus (Stiggins, 2002).

Despite increases in education funding, little improvement has been made and in some cases a drop in the achievement of school children in some countries has occurred (Productivity Commission, 2016; Gonski, 2018). To realize an improvement in education outcomes, a report on the *National Education Evidence Base*, called for the immediate need for refinement and efficiency in the collection and use of education and learning data refinement (Productivity Commission, 2016). The report called for developing the “bottom-up” capability of teachers' use of assessment data and for expanded use of technology in assessment. The present review has highlighted ways in which technology can be used in meaningful ways to enhance assessment processes in classrooms. Technology has the potential to standardize and simplify test administration, to automate test scoring, to create reports that make use of new measures of learning, to customize reports, to deliver reports to a range of stakeholders, to aid in the interpretation of results across different levels of expertise and perspectives, to link assessment to curriculum, be used in ways to inform lesson planning, and used to monitor growth in learning over time.

Countries such as the Commonwealth of Australia and the United States of America are making significant strides in assessing their students at the national level. In the USA there is the National Assessment of Educational Progress (NAEP) (2018) and in Australia there is the National Assessment Program Literacy and Numeracy (NAPLAN) (2018). Despite these feats, there is not yet standardized national assessments in any content area for children of preschool or kindergarten age. Although there are already examples of technology being used for classroom assessments of young learners, there remains a critical need to move education further into the digital age. In the USA, there is a race to fund innovative state assessment systems, such as computer adaptive assessments, that can provide an annual summative determination, validate when students are ready to demonstrate mastery, and allow for differentiated student support based on the individual learning needs (Every Student Succeeds Act, 2015). The Every Student Succeeds Act

(ESSA) shifts its focus on multiple measures rather than a single measure in order to shed further light on the teaching and learning cycles in classroom assessment. The integration of technology in instruction and assessment is still at a nascent stage for both countries, but there is swift transformation taking place.

One of those transformations is the operational definition of technology, a crosscutting concept that touches all aspects of modern life. When children are asked to define technology, they mention computers and electrically powered devices (Pearson and Young, 2002; Lachapelle et al., 2018), Australia and the USA are working to broaden the scope of the general population's definition of technology. In alignment with the National Research Council (2010), the National Assessment Governing Board (2017), and the Australian Curriculum, Assessment and Reporting Authority (ACARA) (2016), technology is defined as all products and processes that solve a problem, need, desire, or opportunity. Australia has gone so far as to declare that technology is the foundation for success in all learning areas [Ministerial Council on Education, Employment Training, and Youth Affairs (MCEETYA), 2008]. Both countries are placing special emphasis on information communication technologies (ICT) because of its crucial role in the workforce. ICT includes the accessing, managing, integrating, and presenting of information [ICT Literacy Panel, 2001; State Education Technology Directors Association, 2003; Ministerial Council on Education, Employment Training, and Youth Affairs (MCEETYA), 2008]. As both countries lead the way in national assessment, the area of technology-based classroom assessment in early childhood education is set to reap the benefits.

In this respect, technology-based assessment has a particularly strong potential to advance AfL practices in the early childhood years, although it can also enhance AoL and diagnostic assessment practices. At present, most of the focus of technology-based assessment in the early childhood years classroom has been from grade 1 and onwards. Assessments are also being increasingly used in earlier years, where the initial motivations have been to screen for school readiness and for benchmarking. Technology-based assessments as a part of AfL practices in kindergarten and preschool needs further research and development. Observations that children as young as 2–3 years can use a tablet to provide a valid assessment of cognitive skills (Twomey et al., 2018) and early literacy skills (Neumann and Neumann, 2018; Neumann et al., 2019) supports this work. Moreover, the development of ways to support the bottom-up capability of teachers to collect AfL and AoL data using digital technology in the classroom is critical. Rigorous testing

of the validity and reliability of test scores produced by digital assessment tools in the classroom will allow children's knowledge to be collected efficiently, cost-effectively, and with accuracy. This approach will increase our knowledge of how technologies are used in assessment, how data can be linked across curriculum areas and across students, and how to represent data for education purposes to achieve individual learning goals and student success.

CONCLUSION

The challenge of making technology-based educational assessments a part of good educational practice can only be met through the joint efforts from a range of stakeholders. It will depend on investments in research to establish a strong evidence base for practice as well as the further research and development of new technology and new uses for existing technology. To be successful, this work will need strong university-industry partnerships and the support of government education departments at the local, state, and national levels. The outcome could see the establishment of digital ecosystems that involve educators, students, and other stakeholders in the design, development, and utilization of practical and utilitarian technology-based assessments. In turn, this could lead to increased efficiency and improved educational outcomes for students across all age levels. The development of committed collaboration among educational researchers, the technology industry, governments, and policy developers are needed to ensure the advantages of technology-based assessment are fully realized.

AUTHOR CONTRIBUTIONS

Idea for review conceived by MN. Research, writing, and revision of manuscript jointly contributed by MN, JA, NE, and DN. Submission by DN.

FUNDING

The research reported here was supported in part by an International Collaboration Award by the College of Behavioral and Community Sciences at the University of South Florida and by the Institute of Education Sciences, U.S. Department of Education, through award number R305A170638. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

REFERENCES

- Abrams, L., Pedulla, J., and Madaus, G. (2003). Views from the classroom: teachers' opinions of statewide testing programs. *Theor. Into Pract.* 42, 18–29. doi: 10.1207/s15430421tip4201_4
- Allinder, R. M. (1995). An examination of the relationship between teacher efficacy and curriculum based measurement and student achievement. *Remed. Spcl. Educ.* 27, 141–152. doi: 10.1177/074193259501600408
- American Academy of Pediatrics (2016). *American Academy of Pediatrics Announces New Recommendations for Children's Media Use*. Retrieved from: <https://www.aap.org/en-us/about-the-aap/aap-press-room/Pages/American-Academy-of-Pediatrics-Announces-New-Recommendations-for-Childrens-Media-Use.aspx>
- Anthony, J. L. (2018). "Dimensionality of English letter knowledge across names, sounds, case, and response modalities," in *Paper Presented at the Annual Meeting of the Society for the Scientific Study of Reading* (Brighton).

- Anthony, J. L., Davis, C., Williams, J. M., and Anthony, T. I. (2014). Preschoolers' oral language abilities: a multilevel examination of dimensionality. *Learn. Individ. Diff.* 35, 56–61. doi: 10.1016/j.lindif.2014.07.004
- Anthony, J. L., Lonigan, C. J., Burgess, S. R., Driscoll, K., Phillips, B. M., and Cantor, B. G. (2002). Structure of preschool phonological sensitivity: overlapping sensitivity to rhyme, words, syllables, and phonemes. *J. Exp. Child Psychol.* 82, 65–92. doi: 10.1006/jecp.2002.2677
- Anthony, J. L., Williams, J. M., Durán, L., Gillam, S., Liang, L., Aghara, R., et al. (2011). Spanish phonological awareness: dimensionality and sequence of development during the preschool and kindergarten years. *J. Educ. Psychol.* 103, 857–876. doi: 10.1037/a0025024
- Anthony, J. L., Williams, J. M., Erazo, N. A., Montroy, J. J., and Cen, W. (2017). *Texas Kindergarten Entry Assessment (TX-KEA): Listening Comprehension Subtest*. Houston, TX: University of Texas Health Science Center at Houston and Texas Education Agency.
- Australian Curriculum, Assessment and Reporting Authority (ACARA) (2013). *National Assessment Program—Literacy and Numeracy*. Retrieved from: <https://www.nap.edu.au/>
- Australian Curriculum, Assessment and Reporting Authority (ACARA) (2016). *The Australian Curriculum: Digital Technologies*. Sydney, NSW: ACARA.
- Australian Institute for Teaching and School Leadership (AITSL) (2011). *Australian Professional Standards for Teachers*. Retrieved from: https://www.qct.edu.au/pdf/QCT_AustProfStandards.pdf
- Aziz, N. A. A., Sin, N. S. M., Batmaz, F., Stone, R., and Chung, P. W. H. (2014). Selection of touch gestures for children's applications: repeated experiment to increase reliability. *Int. J. Adv. Comput. Sci. Appl.* 5, 97–102. doi: 10.14569/IJACSA.2014.050415
- Bailey, A. L., and Heritage, M. (2008). *Formative Assessment for Literacy, Grades K-6: Building Reading and Academic Language Skills Across the Curriculum*. Thousand Oaks, CA: Corwin Press.
- Baird, J., Andrich, D., Hopfenbeck, T. N., and Stobart, G. (2017). Assessment and learning: fields apart? *Assess. Educ.* 24, 317–350. doi: 10.1080/0969594X.2017.1319337
- Becker, D., Bay-Borelli, M., Brinkerhoff, L., Crain, K., Davis, L., Fuhrken, C., et al. (2011). *Top Ten: Transitioning English Language Arts Assessments*. Iowa City, IA: Pearson.
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assess. Educ.* 18, 5–25. doi: 10.1080/0969594X.2010.513678
- Black, P., and Wiliam, D. (1998). Assessment and classroom learning. *Assess. Educ.* 5, 7–74. doi: 10.1080/0969595980050102
- Boyle, A., and Hutchison, D. (2009). Sophisticated tasks in e-assessment: what are they and what are their benefits? *Assess. Eval. Higher Educ.* 34, 305–319. doi: 10.1080/02602930801956034
- Bridgeman, B., Lennon, M. L., and Jackenthal, A. (2003). Effects of screen size, screen resolution, and display rate on computer-based test performance. *Appl. Meas. Educ.* 16, 191–205. doi: 10.1207/S15324818AME1603_2
- Broadfoot, P., and Black, P. (2004). Redefining assessment? The first ten years of Assessment in Education. *Assess. Educ.* 11, 7–26. doi: 10.1080/0969594042000208976
- Bruder, I. (1993). Alternative assessment: putting technology to the test. *Electron. Learn.* 12, 22–23.
- Burnett, C. (2010). Technology and literacy in early childhood educational settings: a review of research. *J. Early Childhood Literacy* 10, 247–270. doi: 10.1177/1468798410372154
- Burns, M., Appleton, J., and Stehouwer, J. (2005). Meta-analytic review of responsiveness-to-intervention research: examining field-based and research-implemented models. *J. Psychoeduc. Assess.* 23, 381–394. doi: 10.1177/073428290520300406
- Busch, T. W., and Reschly, A. L. (2007). Progress monitoring in reading: using curriculum based measurement in a response-to-intervention model. *Assess. Effect. Interv.* 32, 223–230. doi: 10.1177/15345084070320040401
- Carson, K. (2017). Reliability and predictive validity of preschool web-based phonological awareness assessment for identifying school-aged reading difficulty. *Commun. Disord. Q.* 39, 259–269. doi: 10.1177/1525740116686166
- Chang, Y. M., Mullen, L., and Stuve, M. (2005). Are PDAs pedagogically feasible for young children? Examining the age-appropriateness of handhelds in a kindergarten classroom. *Technol. Horizons Educ. J.* 32:40. Available online at: <https://www.learnlib.org/p/77200>
- Chen, J., and Chang, C. (2006). Using computers in early childhood classrooms: teachers' attitudes, skills and practices. *J. Early Childhood Res.* 4, 169–188. doi: 10.1177/1476718X06063535
- Clements, D. H., and Sarama, J. (2003). Young children and technology: what does the research say? *Young Child.* 58, 34–40. Available online at: <https://www.jstor.org/stable/42729004>
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teach. Coll. Rec.* 106, 1047–1085. doi: 10.1111/j.1467-9620.2004.00372.x
- Davey, T., Godwin, J., and Mittelholtz, D. (1997). Developing and scoring an innovative computerized writing assessment. *J. Educ. Meas.* 34, 21–41. doi: 10.1111/j.1745-3984.1997.tb00505.x
- Deluca, C., Luu, K., Sun, Y., and Klinger, D. A. (2012). Assessment for learning in the classroom: barriers to implementation and possibilities for teacher professional learning. *Assess. Matters* 4, 5–29. Available online at: <https://www.nzcer.org.nz/nzcerpress/assessment-matters/articles/assessment-learning-classroom-barriers-implementation-and>
- Deno, S. L. (1985). Curriculum-based measurement: the emerging alternative. *Except. Child.* 52, 219–232. doi: 10.1177/001440298505200303
- Deno, S. L. (1997). "Whether thou goest... Perspectives on progress monitoring," in *Issues in Educating Students With Disabilities*, eds J. W. Lloyd, E. J. Kameenui, and D. Chard (Mahwah, NJ: Lawrence Erlbaum Associates), 77–99.
- Dolan, R. P., Burling, K. S., Rose, D., Beck, R., Murray, E., Strangman, N., et al. (2010). *Universal Design for Computer-Based Testing (UD-CBT) Guidelines*. Iowa City, IA: Pearson.
- Donker, A., and Reitsma, P. (2007). Drag-and-drop errors in young children's use of the mouse. *Interact. Comput.* 19, 257–266. doi: 10.1016/j.intcom.2006.05.008
- Dwyer, C. A. (1998). Assessment and classroom learning theory and practice. *Assess. Educ.* 5, 131–137. doi: 10.1080/0969595980050109
- Ellis, K., and Blashki, K. (2004). Toddler techies: a study of young children's interaction with computers. *Inform. Technol. Childhood Educ. Annu.* 1, 77–96.
- Every Student Succeeds Act (2015). Every Student Succeeds Act of 2015 § 6301, 20, USC § 1177. Retrieved from: <https://nche.ed.gov/every-student-succeeds-act/>
- Falk, T. H., Tam, C., Schellnus, H., and Chau, T. (2011). On the development of a computer-based handwriting assessment tool to objectively quantify handwriting proficiency in children. *Comput. Methods Prog. Biomed.* 104, 102–111. doi: 10.1016/j.cmpb.2010.12.010
- Foorman, B. R., York, M., Santi, K. L., and Francis, D. (2008). Contextual effects on predicting risk for reading difficulties in first and second grade. *Read. Writing* 21, 371–394. doi: 10.1007/s11145-007-9079-5
- Fuchs, L. (2004). The past, present, and future of curriculum-based measurement research. *Schl. Psychol. Rev.* 33, 188–192. Available online at: <https://psycnet.apa.org/record/2004-16860-002>
- Fuchs, L., Deno, S. L., and Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation of pedagogy, student achievement, and student awareness of learning. *Am. Educ. Res. J.* 21, 449–460. doi: 10.3102/00028312021002449
- Fuchs, L. S., Fuchs, D., and Hamlett, C. L. (1989). Effects of alternative goal structures within curriculum-based measurement. *Except. Child.* 55, 429–438. doi: 10.1177/001440298905500506
- Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., et al. (2009). *Assisting Students Struggling with Mathematics: Response to Intervention (RTI) for Elementary and Middle Schools. IES Practice Guide. NCEE 2009–4060*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from: https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_math_pg_042109.pdf
- Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., et al. (2008). *Assisting Students Struggling with Reading: Response to Intervention (RTI) and Multi-Tier Intervention in the Primary Grades. IES Practice Guide. NCEE 2009–4045*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from: https://ies.ed.gov/ncee/wwc/docs/practiceguide/rti_reading_pg_021809.pdf
- Gonski, D. (2018). *Through Growth to Achievement: Report of the Review to Achieve Educational Excellence in Australian Schools*. Commonwealth of Australia. Retrieved from: https://docs.education.gov.au/system/files/doc/other/662684_tgta_accessible_final_0.pdf

- Graney, S. B. and Shinn, M. R. (2005). Effects of reading curriculum-based measurement (R-CBM) teacher feedback in general education classroom. *Schl. Psychol. Rev.* 34, 184–201. Available online at: <https://psycnet.apa.org/record/2005-06945-003>
- Harvard Business Review Analytical Services (2017). *Education in Australia: Transforming for the 2020 economy*. Retrieved from: <https://blogs.msdn.microsoft.com/education/2017/10/30/education-in-australia-transforming-for-the-2020-economy/>
- Hayward, L., and Hedge, N. (2005). Travelling towards change in assessment: policy, practice and research in education. *Assess. Educ.* 12, 55–76. doi: 10.1080/0969594042000333913
- Heritage, M. (2007). Formative assessment: what do teachers need to know and do? *Phi Delta Kappan* 89, 140–145. doi: 10.1177/003172170708900210
- Heritage, M. (2018). Assessment for learning as support for student self-regulation. *Aust. Educ. Res.* 45, 51–63. doi: 10.1007/s13384-018-0261-3
- Hitchcock, C. H., and Noonan, M. J. (2000). Computer-assisted instruction of early academic skills. *Top. Early Childh. Spl. Educ.* 20, 145–158. doi: 10.1177/027112140002000303
- Hodges, D., Eames, C., and Coll, R. K. (2014). Theoretical perspectives on assessment in cooperative education placements. *Asia Pac. J. Cooperat. Educ.* 15, 189–207. Available online at: <https://eric.ed.gov/?id=EJ1113725>
- Hourcade, J. P., Bederson, B. B., Druin, A., and Guimbretière, F. (2004). Differences in pointing task performance between preschool children and adults using mice. *ACM Trans. Comput. Hum. Interact.* 11, 357–386. doi: 10.1145/1035575.1035577
- ICT Literacy Panel (2001). *Digital Transformation: A Framework for ICT Literacy*. Princeton, NJ: Educational Testing Service.
- Jodoin, M. G. (2003). Measurement efficiency of innovative item formats in computer-based testing. *J. Educ. Meas.* 40, 1–15. doi: 10.1111/j.1745-3984.2003.tb01093.x
- Joiner, R., Messer, D., Light, P., and Littleton, K. (1998). It is best to point for young children: a comparison of children's pointing and dragging. *Comput. Hum. Behav.* 14, 513–529. doi: 10.1016/S0747-5632(98)00021-1
- Klein, P. S., Nir-Gal, O., and Darom, E. (2000). The use of computers in kindergarten, with or without adult mediation; effects on children's cognitive performance and behavior. *Comput. Hum. Behav.* 16, 591–608. doi: 10.1016/S0747-5632(00)00027-3
- Koomen, M., and Zoanetti, N. (2018). Strategic planning tools for large-scale technology-based assessments. *Assess. Educ.* 25, 200–223. doi: 10.1080/0969594X.2016.1173013
- Kopriva, R., and Bauman, J. (2008). "Testing for the future: addressing the needs of low English proficient students through using dynamic formats and expanded item types," *Paper presented at the American Educational Research Association* (New York, NY).
- Labbo, D. L., and Reinking, D. (2003). "Computers and early literacy education," in *Handbook of Early Childhood Literacy*, eds N. Hall, J. Larson, and J. Marsh (London: SAGE), 338–354. doi: 10.4135/9781848608207.n28
- Lachapelle, C. P., Cunningham, C. M., and Oh, Y. (2018). What is technology? Development and evaluation of a simple instrument for measuring children's conceptions of technology. *Int. J. Sci. Educ.* 41, 188–209. doi: 10.1080/09500693.2018.1545101
- Landry, S. H., Anthony, J. L., Assel, M. A., Carlo, M., Johnson, U., Montroy, J., et al. (2017). *Texas Kindergarten Entry Assessment Technical Manual*. Houston, TX: University of Texas Health Science Center.
- Lankhshear, C., and Knobel, M. (2003). New technologies in early childhood literacy research: a review of research. *J. Early Childh. Literacy* 3, 59–82. doi: 10.1177/14687984030031003
- Lawless, K. A., and Pellegrino, J. W. (2007). Professional development in integrating technology into teaching and learning: knowns, unknowns, and ways to pursue better questions and answers. *Rev. Educ. Res.* 77, 575–614. doi: 10.3102/0034654307309921
- Looney, A., Cumming, J., Van der Kleij, F. M., and Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assess. Educ.* 25, 1–26. doi: 10.1080/0969594X.2016.1268090
- Lopez, L. M., and Pasquini, R. (2017). Professional controversies between teachers about their summative assessment practices: a tool for building assessment capacity. *Assess. Educ.* 24, 228–249. doi: 10.1080/0969594X.2017.1293001
- Lu, Y., Ottenbreit-Leftwich, A. T., Ding, A., and Glazewski, K. (2017). Experienced iPad-using early childhood teachers: practices in the one-to-one iPad classroom. *Comput. Schl.* 34, 1–15. doi: 10.1080/07380569.2017.1287543
- Luecht, R. M. (1996). Multidimensional computerized adaptive testing in a certification or licensure context. *Appl. Psychol. Meas.* 20, 389–404. doi: 10.1177/014662169602000406
- Martin, W., Strother, S., Beglau, M., Bates, L., Reitzes, T., and McMillan Culp, K. (2010). Connecting instructional technology professional development to teacher and student outcomes. *J. Res. Technol. Educ.* 43, 53–74. doi: 10.1080/15391523.2010.10782561
- Matthews, J., and Seow, P. (2007). Electronic paint: understanding children's representation through their interactions with digital paint. *Int. J. Art Design Educ.* 26, 251–263. doi: 10.1111/j.1476-8070.2007.00536.x
- Ministerial Council on Education, Employment Training, and Youth Affairs (MCEETYA) (2008). *Melbourne Declaration on Educational Goals for Young Australians*. Melbourne, VIC: MCEETYA.
- Mislevy, R. J., Steinberg, L. S., and Almond, R. G. (2003). Focus article: on the structure of educational assessments. *Meas. Interdiscipl. Res. Perspect.* 1, 3–6. doi: 10.1207/S15366359MEA0101_02
- National Assessment Governing Board (2014). *Science framework for the 2015 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board, U.S. Department of Education.
- National Assessment Governing Board (2017). *Reading framework for the 2017 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board, U.S. Department of Education.
- National Assessment of Educational Progress (NAEP) (2018). *Frequently Asked Questions*. Retrieved from: <https://nces.ed.gov/nationsreportcard/about/faqs.aspx>
- National Assessment Program Literacy and Numeracy (NAPLAN) (2018). *When Will NAPLAN Online Start?* Retrieved from: <https://www.nap.edu.au/online-assessment/FAQs>
- National Research Council (2010). *State Assessment Systems: Exploring Best Practices and Innovations: Summary of Two Workshops*. Washington, DC: The National Academies Press.
- Neal, D. (2011). "The design of performance pay in education," in *Handbook of the Economics of Education*, eds E. A. Hanushek, S. Machin, and L. Woessmann (San Diego, CA: North-Holland), 495–550. doi: 10.3386/w16710
- Neumann, M. M. (2018). Using tablets and apps to enhance emergent literacy skills in young children. *Early Childh. Res. Q.* 42, 239–246. doi: 10.1016/j.ecresq.2017.10.006
- Neumann, M. M., and Neumann, D. L. (2018). Validation of a touch screen tablet assessment of literacy skills and a comparison with a traditional paper-based assessment. *Int. J. Res. Method Educ.* 42, 385–398. doi: 10.1080/1743727X.2018.1498078
- Neumann, M. M., Worrall, S., and Neumann, D. L. (2019). Validation of an expressive and receptive tablet assessment of early literacy. *J. Res. Technol. Educ.* 51, 1–16. doi: 10.1080/15391523.2019.1637800
- Nickerson, R. A. (1989). New directions in educational assessment. *Educ. Res.* 18, 3–7. doi: 10.3102/0013189X018009003
- Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., et al. (2015). GraphoGame—A catalyst for multi-level promotion of literacy in diverse contexts. *Front. Psychol.* 6:671. doi: 10.3389/fpsyg.2015.00671
- Olson, A. (2002). Technology solutions for testing. *Schl. Admin.* 59, 20–23. Available online at: <https://eric.ed.gov/?id=EJ642970>
- O'Neill, K., and Folk, V. (1996). "Innovative CBT item formats in a teacher licensing program," in *Annual Meeting of the National Council on Measurement in Education* (New York, NY).
- Parshall, C. G., Becker, K. A., and Pearson, V. U. E. (2008). "Beyond the technology: developing innovative items," in *Bi-Annual Meeting of the International Test Commission* (Manchester).
- Parshall, C. G., Davey, T., and Pashley, P. J. (2000). "Innovative item type for computerized testing," in *Computerized Adaptive Testing: Theory and Practice*, eds W. J. van der Linden and C. A. W. Glas (Norwell, MA: Kluwer), 129–148. doi: 10.1007/0-306-47531-6_7
- Parshall, C. G., Stewart, R., and Ritter, J. (1996). "Innovations: sound, graphics, and Parshall alternative response modes," *Paper Presented at the Annual Meeting of the National Council on Measurement in Education* (New York, NY).

- Pearson, G., and Young, A. T. (2002). *Technically Speaking: Why all Americans Need to Know More About Technology*. Washington, DC: National Academy of Engineering.
- Pellegrino, J. W., and Quellmalz, E. S. (2010). Perspectives on the integration of technology and assessment. *J. Res. Technol. Educ.* 43, 119–134. doi: 10.1080/15391523.2010.10782565
- Penuel, W. R., Fishman, B. J., Yamaguchi, R., and Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *Am. Educ. Res. J.* 44, 921–958. doi: 10.3102/0002831207308221
- Pew Research Center (2015). *The Smartphone Difference*. Retrieved from: <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/>
- Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *J. Technol. Learn. Assess.* 2, 1–43. Available online at: <https://ejournals.bc.edu/index.php/jtla/article/view/1666>
- Pommerich, M., and Burden, T. (2000). “From simulation to application: examinees react to computerized testing,” *Paper Presented at the Annual Meeting of the National Council on Measurement in Education* (New Orleans, LA).
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theor. Pract.* 48, 4–11. doi: 10.1080/00405840802577536
- Productivity Commission (2016). *National Education Evidence Base, Inquiry Report No. 80*. Canberra, ACT. Retrieved from: <https://www.pc.gov.au/inquiries/completed/education-evidence/report/education-evidence-overview.pdf>
- Randel, B., and Clark, T. (2013). “Measuring classroom assessment practices,” in *SAGE Handbook of Research on Classroom Assessment*, ed J. H. McMillan (Thousand Oaks, CA: SAGE), 145–163. doi: 10.4135/9781452218649.n9
- Rankin, J. G. (2016). *Standards for Reporting Data to Educators: What Educational Leaders Should Know and Demand*. New York, NY: Routledge.
- Renaissance Learning (2015). *STAR Reading: Technical Manual*. Wisconsin Rapids, WI: Renaissance Learning.
- Russell, M. (2010). “Technology-aided formative assessment of learning,” in *Handbook of Formative Assessment*, eds H. L. Andrade and G. J. Cizek (New York, NY: Routledge), 125–138.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instruct. Syst.* 18, 119–144. doi: 10.1007/BF00117714
- Salend, S. J. (2009). *Classroom Testing and Assessment for all Students: Beyond Standardization*. Thousand Oaks, CA: Corwin Press. doi: 10.4135/9781483350554
- Scalise, K., and Gifford, B. (2006). Computer-based assessment in e-learning: a framework for constructing “intermediate constraint” questions and tasks for technology platforms. *J. Technol. Learn. Assess.* 4, 1–43. Available online at: <https://ejournals.bc.edu/index.php/jtla/article/view/1653>
- Schmid, R. F., Miodrag, N., and Francesco, N. D. (2008). A human-computer partnership: the tutor/child/computer triangle promoting the acquisition of early literacy skills. *J. Res. Technol. Educ.* 41, 63–84. doi: 10.1080/15391523.2008.10782523
- Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika* 61, 331–354. doi: 10.1007/BF02294343
- Shapiro, E., Angello, L., and Eckert, T. (2004). Has curriculum-based assessment become a staple of school psychology practice? An update and extension of knowledge, use, and attitudes from 1990 to 2000. *Schl. Psychol. Rev.* 33, 249–257. Available online at: <https://psycnet.apa.org/record/2004-16860-007>
- Silseth, K., and Gilje, O. (2019). Multimodal composition and assessment: a sociocultural perspective. *Assess. Educ.* 26, 1–17. doi: 10.1080/0969594X.2017.1297292
- Smarter Balanced Assessment Consortium (2018). *Smarter Balanced Assessment Consortium: 2016-2017 Technical Report*. Oakland, CA: Regents of the University of California.
- State Education Technology Directors Association (2003). *SETDA National Leadership Institute Toolkit*. Burnie, MD: SETDA.
- Stecher, B. M., and Klein, S. P. (1997). The cost of science performance assessments in large-scale testing programs. *Educ. Eval. Policy Anal.* 19, 1–14. doi: 10.3102/01623737019001001
- Stiggins, R. (2010). “Essential formative assessment competencies for teachers and school leaders,” in *Handbook of Formative Assessment*, eds H. L. Andrade and G. J. Cizek (New York, NY: Routledge), 233–250.
- Stiggins, R. J. (2002). Assessment crisis: the absence of assessment for learning. *Phi Delta Kappan* 83, 758–765. doi: 10.1177/003172170208301010
- Twomey, D. M., Wrigley, C., Ahearne, C., Murphy, R., De Haan, M., Marlow, N., et al. (2018). Feasibility of using touch screen technology for early cognitive assessment in children. *Arch. Dis. Childh.* 103, 853–858. doi: 10.1136/archdischild-2017-314010
- Van der Kleij, F. M., Cumming, J. J., and Looney, A. (2018). Policy expectations and support for teacher formative assessment in Australian education reform. *Assess. Educ.* 26, 620–637. doi: 10.1080/0969594X.2017.1374924
- VanDerHeyden, A. M. (2005). Intervention-driven assessment practices in early childhood/early intervention: measuring what is possible rather than what is present. *J. Early Interv.* 28, 28–33. doi: 10.1177/105381510502800104
- VanDerHeyden, A. M., Witt, J.G., and Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *J. Schl. Psychol.* 45, 225–256. doi: 10.1016/j.jsp.2006.11.004
- Verhallen, M., Bus, A., and De Jong, M. (2006) The promise of multimedia stories for kindergarten children at risk. *J. Educ. Psychol.* 98, 410–419. doi: 10.1037/0022-0663.98.2.410
- Wainer, H. (1990). *Computerized Adaptive Testing: A Primer*. Hillsdale, NJ: Lawrence Erlbaum.
- Wainer, H., and Kiely, G. L. (1987). Item clusters and computerized adaptive testing: a case for testlets. *J. Educ. Meas.* 24, 185–201. doi: 10.1111/j.1745-3984.1987.tb00274.x
- Wainer, H., and Lewis, C. (1990). Toward a psychometrics for testlets. *J. Educ. Meas.* 27, 1–14. doi: 10.1111/j.1745-3984.1990.tb00730.x
- Wiliam, D. (2010). “An integrative summary of the research literature and implications for a new theory of formative assessment,” in *Handbook of Formative Assessment*, eds H. L. Andrade and G. J. Cizek (New York, NY: Routledge), 18–40.
- Wiliam, D. (2011). What is assessment for learning? *Stud. Educ. Eval.* 37, 3–14. doi: 10.1016/j.stueduc.2011.03.001
- Wise, S. L. (1996). “A critical analysis of the arguments for and against item review in computerized adaptive testing,” *Paper Presented at the Annual Meeting of the National Council on Measurement in Education* (New York, NY).
- Wise, S. L. (1997). “Examinee issues in CAT,” *Paper Presented at the Annual Meeting of the National Council on Measurement in Education* (Chicago, IL).
- Woloshyn, V. W., Bajovic, M., and Worden, M. M. (2017). Promoting student-centred learning using iPads in a grade 1 classroom: using the digital didactic framework to deconstruct instruction. *Comput. Schl.* 34, 152–167. doi: 10.1080/07380569.2017.1346456

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Neumann, Anthony, Erazo and Neumann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.