Chapter 18
Summative Literacy Assessments and How They Imagine Children:
An International Comparison

Catherine Compton-Lilly
University of South Carolina, USA

Kerryn Dixon
https://orcid.org/0000-0003-4748-6338
University of the Witwatersrand, South Africa

Hilary Janks
University of the Witwatersrand, South Africa

Annette Woods
https://orcid.org/0000-0003-1585-5688
Queensland University of Technology, Australia

ABSTRACT

As an international team of scholars, we have individually and collectively encountered a range of summative and formative assessment practices. Some of these assessment practices have originated from other parts of the world as policy practices increasingly entail global borrowing. We open this chapter with two compelling views of childhood; one places the onus on leading, directing, and controlling children's learning; the other views learning as idiosyncratic, unpredictable, and stunningly contingent on each child's vision of the world. We then introduce readers to a summative assessment associated with three countries (Australia, South Africa, and the United States) to explore how the use of these assessments contributes to the proliferation of particular views of childhood. Finally, we discuss the use of three formative literacy assessments that have gained international attention and present alternative visions of childhood and literacy learning.

DOI: 10.4018/978-1-7998-0323-2.ch018

INTRODUCTION

As scholars from Australia, South Africa, and the United States, we have individually and collectively encountered both summative and formative assessment practices over many years and in many contexts. Some of these assessment practices have global talons originating in other parts of the world as policy practices increasingly entail global policy borrowing. Others are local. Significantly, we are all employed by universities, teach in mid-sized or large urban areas, and engage with communities and schools where the dominant language is English. While each context brings its own set of challenges, policies and resources, we share a commitment to serving children who have been historically underserved by school education. We are particularly interested in how assessments position and treat children who have traditionally been denied access to quality equitable schooling.

As we weigh our concerns and consider the uses of literacy assessments, we propose what we describe as a childhood-literacy-theory nexus, a coming together of views about childhood, literacy, and learning theory that inform and guide how literacy is taught. Conceptions of childhood operating within this nexus include the degree to which childhood is treated as entailing identifiable, predictable, and universal trajectories of growth. Relevant conceptions of literacy reflect whether literacy is conceptualized as skills-based or meaning-based, as decoding or encoding. Finally, our nexus addresses whether learning theories focus on identifiable best practices that are understood as appropriate for all children, regardless of cultural, linguistic, and experiential differences or whether learning processes are conceptualized as socio-culturally and individually unique. These theoretical conceptions of learning have significant implications for teachers who can be treated as implementers of existing programs or experts who design instruction for children.

Across this nexus, we have become increasingly concerned about the flow and circulation of largescale summative literacy assessments; how different constructions of literacy and childhood are interwoven and how these different constructions inform these literacy assessments. Our argument is that these assessments have a limited approach to literacy, a narrow view of childhood and detrimental effects on children, teachers and early childhood education more generally. In this chapter, we explore dimensions of a childhood-literacy-theory nexus. We begin by examining the mutually constitutive nature of theories of early literacy and theories of childhood. Then we present three largescale summative assessments from three different international contexts, Australia, South Africa and the United States and their theories of literacy and childhood. Having examined what they achieve and in what ways they are problematic, we turn to formative approaches to assessment as an alternative.

Theories of Childhood, Early Literacy and Assessment

Two longstanding and compelling views of childhood present different nexical formations related to childhood, literacy, and theories. Both situate children as capable of learning and knowing; however, one places the onus on teachers to lead, direct, and control children's learning. According to this way of thinking, the goal of assessment is to measure and evaluate learning in relation to linear, and pre-determined learning progressions (Apple, 2011; Ravitch, 2013). When applied to literacy, this view highlights the mastery of specific and often isolated skills over time (Hammill & Swanson, 2006; National Reading Panel, 2000; Wyse & Goswami, 2008; Wyse & Styles, 2007) and has been linked to neoliberal views of literacy education (Edmondson, 2001; Hibbert, Heydon, & Rich, 2008; Rogers, Mosley, & Folkes, 2009).

The other view (see for example Calkins, 1986; Clay, 1991; Gandini, 1993; Goodman, 1996 Montessori, 2013; Read 1986/2018; Teale & Sulzby, 1986) treats learning as idiosyncratic, unpredictable, and stunningly contingent on each child's vision of the world, particular interests and past experiences, as well as the values, languages, social and cultural practices of the communities in which children operate. This view recognizes that literacy learning occurs through sustained exposure to and meaningful interaction with text that is deeply embedded in cultural and familial practices and perspectives. While this view does not preclude attention being paid to particular skills and strategies, it is linked to child-centered views of learning and childhood.

While we argue that any given assessment could be used either summatively or formatively, we note the current international privileging of summative uses of assessments. We maintain that the ways assessments are used – summatively or formatively – present different notions of childhood, literacy, and learning theories that have implications for how assessment is conducted and how assessment data is valued and used. In what follows we briefly review each perspective and reveal critical epistemological differences.

Researchers have claimed differing purposes for summative and formative assessments (Afflerback, 2016; Dixon & Worrell, 2016; Dolin, Black, Harlen, & Tiberhein, 2017; Wixson, 2017). For example, Wixson (2017) argues that formative assessments are used "on an ongoing basis in daily instruction" (p. 78) to ascertain the appropriateness of particular instructional methods and to help the teacher review instructional strategies as needed, while summative assessments operate a screening tools or to "measure progress" (p. 78). Other scholars (Dixon & Worrell, 2016; Dolin, Black, Harlen, & Tiberhein, 2017; Wixson, 2017) describe summative and formative assessments as complementary and/or as operating on a continuum that ranges from informal formative to formal summative. Others (Black & William, 2009, 2018) question whether assessments are inherently either summative or formative assessments, suggesting instead that the differences between assessments result from the inferences made by teachers about the children they are assessing. When such inferences "relate to the status of the child, or about their future potential, then the assessment is functioning summatively" (p. 553). On the other hand, where teachers make inferences related to how best to "help the student learn, then the assessment is functioning formatively" (Black & Wiliam, 2009, p. 553). Discerning the differences between the two assessments can be helpful (Shearer & Klenowski, 2019). Summative uses of assessments focus on the end point of learning usually after intentional teaching sequences, while formative uses of assessments focuses on monitoring ongoing progress. While recognizing these distinctions, in this chapter, we are primarily concerned with the conceptions of childhood and literacy learning that accompany the use of assessments.

In 2006, Cochran-Smith and Lytle presented a critique of *No Child Left Behind* (U.S. Department of Education, 2004) that highlighted three images of teachers and education implicit in that U.S government report: images of knowledge, images of teachers and teaching, and images of teacher learning. They argued that these images were "flawed – linear, remarkably narrow and based on a technical transmission model of teaching, learning, and teacher training that was rejected more than two decades ago" (p. 669). According to Cochran-Smith and Lytle inherent in this model is a vision of teaching as "fundamentally technical, instrumental, certain and decontextualized" (p. 681) rather than "complex, deliberative, inherently uncertain, contextual, social, cultural, and political" (p. 682). In the NCLB policy initiatives, teacher expertise was negated in favor of assessments and scores that are assumed to translate simply into best practices. In this chapter, we make a similar argument focused on images of childhood and children's learning.

Theoretical differences related to children and learning affect the international flow of assessments – used for both summative and formative purposes. Here we specifically ask what circulates, for what purposes, and what do these global flows mean for children and their becoming literate. We then introduce readers to assessments associated with the three countries in which we work, as well as policies and purposes surrounding their use, in order to explore how these assessments contribute to the proliferation of particular views of literacy and childhood. Summative literacy assessments we address include The *National Assessment Program – Literacy and Numeracy* (NAPLAN) from Australia and *Progress in International Reading Assessment Study* (PIRLS) used in South Africa. These are examples of assessments used by teachers, schools and systems to collect data about children and infer their capacities in a summative fashion across numerous contexts in current times. We then consider the *Early Grade Reading Assessment* (EGRA), an assessment created in the United States of America and implemented internationally usually toward monitoring or formative purposes. Thus, we present the examples of large-escale benchmarked assessments - that generally function summatively to evaluate educational systems related to literacy – and often present neoliberal notions of childhood and learning. We worry that these assessment systems do not provide nuanced understandings of individual children's literacy learning.

Neoliberal Views of Childhood

Neoliberal views of childhood and literacy learning place the onus on best practices that correlate with the mastery of foundational reading skills in phonemic awareness, phonics, vocabulary, fluency, and comprehension (National Reading Panel Report, 2000). These foundational areas of knowledge are identified as the "pillars" that uphold the emerging reading processes of young readers (Hammill & Swanson, 2006; National Reading Panel, 2000; Wyse & Goswami, 2008; Wyse & Styles, 2007). Foundational skills are assumed to be universal and sequential; we argue that they reflect narrow, linear, and pre-determined learning progressions (see critiques, Apple, 2011; Ravitch, 2013). Thus, the role of teachers is to lead, direct, and control children's learning through sequences of activities that have been identified as foundational to learning to read. Teacher expertise involves fidelity to identified instructional techniques, and assessment involves the documenting of achievement relative to each of the "pillars." These pillars, once mastered, are assumed to translate into accomplished reading.

In Stevens' review of *Reading First* (2003), she refers to this as a "cultural model of reading" that presents a particular view of readers, reading programs, and teachers. In regard to readers, she argues, "What is missing from this narrow view of the reader is the possibility, the necessity of explicitly confronting the historically situated differences among languages and developing students' awareness of both" (p. 665). Stevens and others (Cochran-Smith & Lytle, 2006; Darling-Hammond, 2007) share concerns about the denigration of teacher expertise and the need for teachers to focus on children rather than programs and the complexities that can accompany learning to read.

Due to their avowed commitment to documenting the learning of identifiable and assumedly universal literacy skills, the assessment techniques that accompany these models tend to focus on summative end goals to provide information about instructional programs, schools, and/or systems. However, these assessments are sometimes used to assess individual children relative to their peers, and more recently to judge teachers and even teacher education training programs. Murphy (2015) argues that the recent proliferation of neoliberal and skills-based approaches has resulted in the assessment of young children becoming "inscriptive rather than prescriptive" with the propensity to "regulate rather than open up possibilities for learning" (p. 26).

Flewitt and Roberts-Holmes (2015) worry that neoliberal, best practice approaches result in situations where "teachers' locally based, complex, subtle, and professional judgments about children's [literacy] learning as an ongoing process become reduced to simplistic, crude sets of normalized data, and any notion of children's holistic learning is replaced by statistics" (p. 106). To make their point, they focus on the *Phonics Screening Check*, which uses pseudo-words to assess children's phonic knowledge and decoding strategies in the UK. The idea of assessing children's reading by asking them to sound out letter combinations that are not in fact able to be comprehended has been critiqued in the United Kingdom and also in contexts such as Australia where such approaches have been borrowed (see for example Clark 2017).

Child-Centered Views of Childhood

A competing perspective of children, views children's learning as idiosyncratic, unpredictable, and stunningly contingent on each child's vision of the world, particular interests and experiences, including the practices of the communities in which they participate. Children are viewed as engaged, agentive, and participating in literacy processes. These views are grounded in observations of individual children as they engage with learning activities and materials (Gandini, 1993; Montessori, 2013), reading activities (Clay, 1991; Goodman, 1996) and writing practices (Calkins, 1986; Read 1986/2018; Teale & Sulzby, 1986). While there is a recognition of general developmental trends and milestones, adherents of this view recognize that children take different paths toward becoming literate. Thus, teacher expertise involves careful observation of individual learners and the crafting of learning experiences that build upon what children know and can do while working to develop competencies that are not yet established. Teacher expertise then involves careful and informed decision-making about particular children rather than the implementation of scripts or assumedly universal best practices. Due to their commitment to documenting the qualitative learning of individual children, assessments supported by these models or approaches tend be formative in nature, aimed at providing information about what particular children know and can do with text; the results or inferences made from these assessments (Black and Wiliam 2018) are then used to design instruction for individual children or groups of children with similar reading and writing patterns and practices. Reflecting these child-centered views, Fontaine and Luttrell (2015) challenge educators to attend to children's multimodal meaning making to avoid the "reduction of children's knowing, being, and doing into its measurable parts" (p. 49).

Unlike neoliberal understandings that support assessments that focus on children's learning of assumedly universal and sequential reading skills, child-centered approaches are uniquely useful for inviting, appreciating and honoring unexpected responses from children. When reading a particular text, a child whose first language is Mandarin Chinese might produce different pronunciations, apply different syntactical structures, and make different meanings from English text than children whose first language is English. Processing differences can be related to past experiences with books and oral storytelling, language, cultural practices and understandings, interactional styles, and home literacy practices. They are not considered markers of problems with reading; instead they are reflections of particular ways of learning to make sense of written text. Applying assessment practices to formative ends, because this would usually involve close observation of children as they engage with literacy, is uniquely suited to document processing differences and to treat them as clues into a child's reading process and as a lens toward understanding the strengths that children bring to text. As part of a child-centered view of literacy, these differences account for many of the idiosyncratic, unpredictable, and socially and culturally informed differences that are observed when assessing children formatively.

These theoretical differences have significant implications for assessment practices across the childhood-literacy-theory nexus. Assessments can be used to measure dimensions of reading processes (phonemic awareness, phonics, vocabulary, fluency, and comprehension) (National Reading Panel Report, 2000) or focus on children's understandings and strategic actions as they engage in making meaning through the receptive and productive dimensions of engaging with text (Clay, 1991; Goodman, 1996, Woods & Comber, 2019).

Considering the International Circulation of Summative Literacy Assessments

As international scholars who have collaborated on various projects, spoken at the same conferences, and socialized in international spaces, our conversations have often turned to concerns about the effects of the international circulation of neoliberal and standardized literacy assessments. We worry specifically about the effects of the summative use of assessments on teacher expertise, instructional practices and the accompanying visions of children, childhood, and learning. In the section below, we identify literacy assessment and accountability policy trends that support notions of assessment targeted toward providing summative assessment data. We first demonstrate how these policy logics might play out in local contexts of Australia, South Africa and United States, and then move to unpack the international roots of these locally administered assessments and significant effects of these international connections on global literacy education. We focus on a small set of assessments that have been used in summative ways to evaluate and compare schools, school systems, nations, and children. These assessments have operated at the nexus of childhood-literacy-theory in ways that privileges neoliberal thinking and technicist views of literacy learning.

We then turn our attention to a small set of formative assessments that also have an international presence. These assessments are generally used to monitor children's progress over time to inform planning and teaching decisions. Privileging a child-centered and sociocultural view of learning, these assessments present an alternative childhood-literacy-theory nexus that not only recognizes children as unique learners, but also tries to accommodate the cultural knowledge of children. An important point to note is that formative assessments used with individual children require minimal materials. While these assessments can have significant financial gains for authors or corporations, they do not require associated scoring mechanisms, test preparation materials, or online data organizational systems, which have created lucrative financial markets for commercial publishers. Instead they are built on assumptions about the importance of teacher expertise, and the need for assessment and teaching that is individually and culturally responsive.

In the sections below, we identify summative assessment practices with international roots or tendrils. We describe each assessment, discuss the official purpose for the assessment, discuss how they have been used, and describe the international roots and/or circulation of that assessment. Finally, we look across the three assessments, identify shared assessment practices and accompanying assumptions about children and literacy learning.

Australia: NAPLAN

Like many other nations, Australia has had a more than two-decade long obsession with assessment, accountability standards, and neoliberal reform. During this time, the Australian system witnessed the introduction of state-based assessments to enable comparative benchmarking of state and territory sys-

tems. More recently a national testing program has been implemented and a public website that reports school-by-school results on national tests now provides ready access to this comparative data and enables the construction of comparative league tables. As part of the same suite of policy moves Australia's first national curriculum was recently designed with implementation beginning in 2010 and continuing. At the same time, teacher standards, a literacy and numeracy test for preservice teachers, and recent moves toward evaluating teachers and teacher educators have been introduced. Unfortunately, the implementation of these policies has led to a standardization vision of children that neither reflects nor honors the diverse populations in Australia nor the spaces in which they live.

A centerpiece of this mandated assessment regime is the National Assessment Program – Literacy and Numeracy (NAPLAN). NAPLAN was introduced in 2008 to test children in years three, five, seven, and nine using a suite of tests across four domains: reading, writing, language conventions, and numeracy. Language conventions include isolated skills (i.e., spelling, grammar, punctuation). Like many of the State literacy tests used in the United States, the tests include multiple choice, short answer, and essay responses. Children take the tests each May. In the weeks and months prior to the testing, significant learning time is allocated to practicing test-taking skills. The tests are scored by a combination of computer assisted marking and marking by trained teachers using established criteria,

Online versions of NAPLAN have been piloted since 2016, and starting in 2018 schools across Australia have had the option of opting-in to administer the NAPLAN online rather than as a paper-based test. Challenges that accompany making the assessment available in remote areas, continued difficulties with connectivity, and continuing inequity in access to technology has slowed the proposed adoption of the online NAPLAN suite and the removal of the print-based version from the options available to schools and systems.

Official Purposes of the Assessment

NAPLAN is designed to "identify whether all students have the literacy and numeracy skills that provide a critical foundation for their learning" (Australian Curriculum, Assessment and Reporting Authority (ACARA, nd). Initially the tests were billed as a "health check" for the nation's education system. NAPLAN is a summative assessment measure and all students in targeted years complete the same tests. However, the tests change each year, which requires that systems and schools implement high levels of security to ensure the test's fidelity. As the tests move online for some students, some children compose written texts with paper and pencil and others do the same on a computer or digital device.

The results of the tests are reported each year in the National Report of Achievement on Reading, Writing, Language Conventions and Numeracy, which reports results on reading, writing, spelling, grammar and punctuation, and numeracy a well as participation rates. Results are reported for Australia and individual states/territories. State results are disaggregated according to gender, Indigeneity, geolocation, parental education level, and parental occupation. As in the United States, the assessment outcomes are reported on public websites that provide school level data for each school in Australia as well as comparisons to demographically similar schools. Systems, schools, and parents are also provided with individual student data. There have been recent efforts to use these data to evaluate teachers and the teacher education institutions. As the uses of these data continue to expand, the reliability and validity of the tests becomes more and more precarious.

International Roots and Circulation

The ACARA logo appears on all NAPLAN tests and materials. This gives the impression that the tests are written by ACARA personnel - employees of the Australian Government. However, ACARA staff generally 'oversee' the process and at least some items, and some aspects of test development, review, and piloting are outsourced to test development companies. In addition, in some States/Territories the printing and delivery of tests, scoring, and the preparation of school and parent reports are also outsourced. ACARA provides limited information about test developers. Officially, printing, delivery, marking and reporting are responsibilities of State/Territorial governments and not ACARA; however, details related to who is doing what across Australia are difficult to find. What is known is that edu-businesses are involved in this Federally mandated assessment program, and that at least some of the work is being carried out by international publishing houses including Pearson. For example, in an analysis of the 2012 NAPLAN design process, Hogan (2016) found that all Australian States except Queensland engaged Pearson to print and distribute the tests, and at least three systems also engaged Pearson to mark the tests and prepare data for reporting. Hogan reported that ACARA contracted several businesses including the Australian Council for Educational Research and Pearson to complete tasks including item development, item piloting, and analyzing and reporting results. Pearson (nd) reported that they supported curriculum agencies to "handle the huge operation that is involved in running" (no page) NAPLAN each year, however the sheer scale of Pearson's involvement, the numbers of large businesses involved, and the amount of money used to outsource these tasks remains difficult to track.

As edu-businesses become increasingly involved in core educational tasks including assessment, pedagogy and curriculum, they are also significantly affecting public policy. Not only are local communities being pushed out of decision-making forums while profit-driven businesses are invited in, but the assessment of children in local schools is being influenced by international priorities and practices that do not necessarily support child-centered notions of literacy learning. Policies that support standardization, accountability, and highly defined curriculum and assessment in Australia, are part of a global push toward standardizing curriculum, pedagogy and assessment to avowedly ensure student learning and teacher quality. The NAPLAN and its accompanying neoliberal policies and practices envision children's literacy learning as aimed at mastery of isolated skills and following a universal and predictable path that can be tracked through summative assessment. The focus is to enforce school and teacher accountability and standardization and to manage ways to compare schools across Australia. Children's test scores are merely a tool in this system.

South Africa: The PIRLS

While NAPLAN is supposedly a locally produced, tailored assessment for the education context it is used within, the Progress in International Reading Literacy Study is an international systemic evaluation of literacy in children's primary language or the language of instruction administered in the 4th Grade. It has been administered every five years since 2001 by the International Association for the Evaluation of Educational Achievement (IEA). It aims to provide internationally comparative data on children's reading achievement.

PIRLS assesses children's reading achievement in their home language or the language of learning and teaching. The test is levelled for children in Grade 4. It is systemic in that it is administered to children across a country, state or territory. Children are tested on their comprehension of literary and informa-

tion texts at different cognitive levels: information retrieval; interpretation and integration of ideas and information; straightforward inferencing; examination and evaluation of content language and textual elements (Mullis, Martin, Foy, & Drucker, 2012). The questions are scored according to whether children's comprehension is complete, partial or absent. The PIRLS test consists of 2 passages selected from an available set of 12, six of which are literary and six information texts. Children answer 13 to 15 multiple choice or short answer questions based on two of the passages. The test takes 30 minutes to complete.

The test is accompanied by questionnaires completed by parents, teachers, school principals and children to identify contextual factors that are likely to affect students' achievement. This includes information on children's "attitudes towards reading and their reading achievement" (Mullis, Martin, Foy, & Drucker, 2012, p. 202), their access to literacy resources, access to libraries, and home and school experiences (https://nces.ed.gov/surveys/pirls/).

Initially, the PIRLS results for low-scoring countries were significantly below the international benchmarks making them too low to enable international comparisons. Thus, a simpler but scalable test, PIRLS Literacy (formerly the Pre-PIRLS), was developed. For example, in 2016, grade four children in five countries (South Africa, Morocco, Egypt, Kuwait, Iran) took the PIRLS literacy test. Denmark's grade three learners also took the test and outperformed the other five countries. After these tests had been scaled, South Africa remained the lowest scoring country with 78% of its grade four children being reported as unable to read and understand literary or information texts across all eleven official languages[1].

Official Purposes of the Assessment

The purpose of PIRLS is to provide internationally comparative data on children's reading achievement and 'provide the best policy relevant information about how to improve teaching and learning and to help young students become accomplished and self-sufficient readers' (PIRLS 2016, p. 4). It is a systemic test that allows for an analysis of national and international trends across time.

Different countries have their own purposes for administering PIRLS. South Africa chose to participate in PIRLS due to poor results on the Trends in Mathematics and Science Study (TIMMS) and the recognition that children's levels of literacy affected their mathematics and science performance. The decision was also based on the fact that PIRLS is also the only large scale, independently administered, benchmarked test that purports to provide reliable systemic information to enable comparisons between provinces, languages, reading purposes and students' gender.

International Roots and Circulation

The International Association for the Evaluation of Educational Achievement (IEA) oversees the PIRLS and TIMMS tests. It is located in Boston and has close links to IEA offices in Amsterdam, Hamburg as well as to Statistics Canada. The USA, Canada, England and Australia are involved in developing test items, sampling activities and monitoring results.

Although PIRLS is an international cooperative of research institutions and government agencies the test was developed by English speaking countries in the political north. In 2016, 50 countries - not including countries tested for benchmarking purposes - participated in PIRLS this is a significant increase from the 36 educational systems that participated in 2001.

Figure 1 shows that PIRLS is generally administered in the political north.

Figure 1. Map of countries participating in PIRLS 2016

The test is developed in English and translated into other languages with the assumption that literacy achievement is comparable across languages. The 2016 South African PIRLS report (Howie, et al., 2017) describes concerns about whether translation affects content and acknowledges the difficulty of translating words from English to African languages when there are no equivalent terms in both languages. Two issues are raised. First, if the language has no equivalent word what does this suggest about the cultural salience of the concept? Second, focusing on content ignores linguistic comparability across non-cognate languages. For example, African languages in South Africa have transparent orthographies but are morphologically complex. As a result, what may be simple to read in an analytic language like English, may be complex in agglutinating languages such as the African languages. We cannot illustrate this comparison using items from PIRLS because the test items in the public domain are only available in English. However, we can demonstrate this complexity by referring to another systemic evaluation. The Annual National Assessments (ANAs) were administered to all children in grades one, three, six, and nine in South Africa in 2014. The tests were written in English and then translated into African languages. The following is an extract from a grade 3 English and isiZulu narrative text:

English:

Ant ran to the river and began to drink. Suddenly a strong wind blew Ant into the water.

isiZulu:

uNtuthwane wagijima waya emfuleni waqala ukuphuza amanzi. Ngokuphazima kweso, kwaqhamuka umoya onamandla wavunguza wadudulela uNtuthwane emazini.

Although there are fewer words in the passage in isiZulu because of its agglutinating nature the average word length is 3.2 letters in English compared to isiZulu's 7.9. The isiZulu text has far more multisyllabic words and letter strings than the English text (see Table 1). While ANAs and PIRLS purport to test comprehension for early readers across languages, they assume that decoding is in place, and that it is a similar process for all children in all languages.

Research shows that even proficient readers take longer to master the orthography for isiZulu in comparison to English, and both its syllabic complexity and the visually similar nature of letter strings

Table 1. Comparison of word complexity of a translated extract

	English	isiZulu
Words	18	16
Average word length	3.2 letters	7.9 letters
Syllables	23	65
Digraphs	2	10

impact automaticity (Land, 2015). Alongside this is the fact that there are no standardized measures of reading proficiency in African languages (van Rooy & Pretorius, 2013) and fewer texts available to children. The choice of language variety for the test is an additional complication that is not limited to South Africa. As another example, children in Egypt and Morocco were tested in school Arabic which is not the variety of Arabic spoken in their communities.

The test also assumes comparability across contexts. While the grade four slump is an international phenomenon (i.e., Chall & Jacobs, 1983), it may be exacerbated by language policies in different countries. In South Africa, Grade 4 is the year in which many children switch from their home language to English as the language of learning and instruction. This is significant because in many other countries the language of instruction remains constant. Although PIRLS questionnaires provide useful information about local socio-cultural and educational contexts, that information is treated as background information rather than as key to complicating claims of international comparability.

It is clear from the results that poorly resourced schools in countries with high poverty do not do well or improve between test cycles which is extremely demoralizing for teachers. And although the test focuses on comprehension, policy makers continue to view teaching decontextualized phonics as the best way to improve results. In fact, England's School's Minister, identified synthetic phonics as the reason for England's improved results on PIRLS (Buckingham, 2017). Whether this is defendable or not in the UK, because African languages have transparent orthographies, phonics alone cannot account for poor results in the South African context. Attention must be paid to what makes reading African languages difficult, and to reading practices that go beyond conceptualizing comprehension as only fluent decoding.

Like NAPLAN, PIRLS is implemented to allow national and international comparisons and track the success of schools and school systems. Designed for an international audience PIRLS cannot put children's local experiences and cultural practices at the center. Universal visions of childhood and children's learning fail to honor the diversity of children's background knowledge equitably.

United States: EGRA

The Early Grade Reading Assessment (EGRA) was developed in 2006 by the Research International Triangle for USAID. It is described by Grove and Wetterberg (2011) as "one tool used to measure students' progress toward learning to read" (p. 2). Their report references DIBELS – an assessment that rose to popularity in the United States in conjunction with the *Reading First* - as one of several approaches that informed the development of EGRA. Critics maintain that DIBELS, with its emphasis on skills identified in the National Reading Panel Report (phonemic awareness, phonics, vocabulary, fluency, comprehension), served as the "model" (Hoffman, 2012, p. 342) or "foundation" (Bartlett, Dowd, & Jonason, 2014, p. 309) for EGRA.

This relationship is significant. First, the modeling of EGRA on DIBELS highlights the commitment that the developers had to a neoliberal and skills-based definition of literacy. The DIBELS examination was created at the University of Oregon by Good and Kaminski (1996). It entails a series of one-minute assessments that measure the acquisition of supposedly discrete literacy skills, including phonemic awareness, alphabetic principle, accuracy and fluency, vocabulary, and comprehension. These skills are posited as contributing to the development of effective reading.

Significantly, the validity of these claims has been challenged on multiple fronts. First, empirical research has generally failed to support adherence to methods that over-rely on systematic phonics and decoding-based approaches to learning to read (Foorman & Connor, 2011; Wyse & Goswami, 2008; Wyse & Styles, 2007). Second, based on their review of existing research, the National Reading Panel Report warned about overly skills-based approaches (Yatvin, Weaver, & Garan, 2003). Finally, empirical research has documented disappointing results for *Reading First* (Games, Bloom, Kemple, & Jacob, 2008; Jackson, McCoy, Pistorno, Wilkinson, Burghardt, Clark, 2007). As Hoffman (2012) and Bartlett, Dowd, and Jonason (2014) argue, the problematic claims related to the efficacy of overly skills-based approaches in the United States invite questions about exporting EGRA - a modification of DIBELS - to nations that struggle already to provide effective literacy learning opportunities to children.

Official Purposes of the Assessment

Echoing the rationale for the creation of Pre-PIRLS, advocates of the international use of EGRA cite the need for assessments that are sensitive to early levels of literacy learning and can track growth for emergent readers. Dubeck and Gove (2014) support EGRA as an assessment that serves "as a baseline of early reading acquisition" (pp. 315-316), as a "guide to the content that is included in an instructional programme" (p. 316), and as a means to "evaluate programmes" (p. 316). Bartlett, Dowd, and Jonason (2014) challenge the scope of these claims suggesting that EGRA also fulfills more utilitarian functions:

EGRA was used by RTI [Research Triangle Institute] and partner organizations as a policy mechanism in multiple sites to demonstrate to people within ministries of education just how little students were learning in the early grades. Assessing skills enabled strong advocacy campaigns and fueled industry-wide shift from talking about learning and quality to measuring reading and questioning quality in the face of rock-bottom scores. (p. 310)

In short, by identifying the need for literacy initiatives in international spaces, EGRA creates a market for investment in international reading education, convincing donors and NGOs of not only the need for financial investment but also the assurance that assessment outcomes will be monitored and reported.

International Roots and Circulation

EGRA is a version of DIBELS that has been adapted for use in African nations where international NGOs use it to track children's reading progress. As in South Africa, English is not the heritage language of most African children. In different parts of Africa many NGO funded schools teach children to read in English. Thus, the EGRA is often administered in English. Language of assessment – whether children are being tested with a translated assessment in their first language or in their second language raises significant concerns. As described above in reference to the PIRLS, differences among languages can

compromise the validity and comparability of scores. For example, Bartlett, Dowd, and Jonason (2014) note that how children respond to phonemic awareness tasks is different for children learning to read through syllabic methods. In addition, the transparency and complexity of letter/sound relationships can affect how and to what extent phonetically-based instruction and assessment are appropriate.

Writing with a focus on the appropriateness of English-based assessments in international spaces, Hoffman (2012) reviews available studies related to administering EGRA in African countries. Not only were these studies not published in peer-reviewed journals, they also lacked compelling findings pointing to the usefulness of EGRA in international settings. In particular, "floor effects" (p. 346) resulted in large numbers of children scoring too low on the test to enable results to be useful.

Bartlett and Vavrus (2016) argue that problems related to the proliferation of EGRA extend beyond problems with the assessment itself. They identify additional issues: 1) the narrowing of the definition of literacy despite international differences in languages, scripts, and notational systems, 2) the problematic reduction of a focus on vast learning opportunities to a narrow focus on reading acquisition, 3) the shift from a focus on learning opportunities to reading performance, and 4) the accompanying rise of an "audit culture" (p. 562) that places little attention on local linguistic, historical, and cultural contexts. These concerns point to issues related to conceptualizations of childhood and childhood learning. In short, EGRA assumes that literacy learning can be measured by a single set of roughly translated assessments regardless of language, culture, or local context.

These assessments accompany instructional programs that strive to create the most learning for the least cost. By measuring a narrow set of literacy skills, these assessments and programs risk focusing on individual performance rather than the creation of rich and engaging learning opportunities and investment in opportunities for children to learn. Measuring learning is not the same as providing rich, culturally, and linguistically responsive learning opportunities. As a result, learning is treated as a universal process that can be best served through scientifically identified best practices. The diversity, creativity, and unique nature of individual children is denied and learning is reduced to the mastery of a sequences of skills without attention to the language, culture, and experiences.

Problems with Using Largescale Summative Assessments

NAPLAN, PIRLS and EGRA present visons of a childhood-literacy-theory nexus that has been designed to enable comparisons across schools, and within and across countries. Ostensibly, the tests will identify unsuccessful schools and provide information to improve those schools and support their students; however, in reality the tests identify broad areas of concern and fail to provide nuanced information that could truly drive appropriate instruction for individual children or even cohorts of children. We identify five problems associated with the use of largescale summative assessments.

- Theoretically these assessments are grounded on a combination of neoliberal views of childhood and skills-based, narrow, and mechanical views of literacy that enable and invite data–driven scripted and programmatic instruction in elementary classrooms around the globe.
- These assessments have been systematically used to invoke and respond to universal conceptions of childhood, literacy and language(s); they are better suited to enable the diagnosis of particular systems rather than individual children and can only identify general areas of individual need in relation to the skills that have been tested.

- The summative assessments presented above bring a strong monolingual bias and ignore the linguistic and literacy repertoires of multilingual children.
- These assessments have been used in ways that undermine and negate teachers' professional knowledge and judgement.
- Finally, these assessments are driven by big business and the investment of capital rather than the interests of children.

These five problems not only reveal how these assessments operate at the nexus of particular views of childhood, literacy learning, and literacy theorization, but also reveal the need for alternative assessments.

In many cases, summative assessments are accompanied by lists of activities that target generalized literacy skills (i.e., phonemic awareness, phonics, fluency). However, these rough assessments do not provide the degree of nuance that can be gained by a professional teacher watching children closely as they engage with literacy tasks. In fact, summative assessments require teachers to lead, direct, and control children's learning, creating a washback effect that affects what it is that teachers notice and believe about learners. This has resulted in the implementation of generic sets of strategies addressing general types of skills without attention the specific challenges faced by particular children, nor their capacities. Thus, children's literacy learning is treated as a universal process of learning, a defined set of essential skills and processes, and as being able to be supported through generic sets of instructional strategies. This is despite counter-evidence available over decades (Foorman & Connor, 2011; Games, Bloom, Kemple, & Jacob, 2008; Jackson, McCoy, Pistorno, Wilkinson, Burghardt, Clark, 2007).

Investing in narrow and assumedly universal models of literacy and literacy learning is advantageous to large international fiscal entities including corporations, large non-profit institutions, and governmental initiatives. All three of the assessments share a focus on measurable discrete skills. NAPLAN assesses children's learning of reading, writing, and language conventions including spelling, grammar and punctuation. PIRLS assesses comprehension skills, including information retrieval, inferencing, and interpretation. EGRA focuses on the five pillars of reading (e.g., phonemic awareness, phonics, vocabulary, fluency, comprehension). Thus, these assessments not only enable national comparisons within their respective countries - with PIRLS and EGRA also enabling international comparisons, but these assessments also pave the ways for test preparation materials and post-test interventions that can be marketed to serve children who score poorly on particular aspects of literacy, and increasingly adjunct materials for parents to use and profit based tutoring schemes that present their services as essential to success.

However, these instructional materials and programs fail to recognize the diversity of the children served, both in terms of what they understand about literacy and in regard to social, cultural and linguistic backgrounds of the children and the diverse geographical locations that they live. While the NAPLAN is only available in Standard Australian English (SAE), attempts to translate PIRLS and EGRA into local languages have often been clumsy and have relied on direct translation rather than linguistically and culturally sensitive rewritings. NAPLAN has been critiqued for ignoring the language and cultural capacities of children whose first language may not be SAE.

Perhaps even more concerning are the proposed and emerging uses of these assessments. All three of the assessments are either being used or are being considered as a means to inform teacher decision-making. However, the summative and/or timed nature of the assessments fail to provide nuanced information that fully recognizes the competencies of children. Instructional implications based on these summative assessments are often lists of instructional activities that correlate to broad areas of weakness. In addition, there is an emerging trend – in the case of NAPLAN and PIRLS – towards administering

these tests online further removing the teacher from the assessment context and significantly reducing opportunities for teachers to observe children as they engage with the assessments.

Involvement of large corporations and organizations in the development, administration and scoring of these assessments alongside the relative neglect of linguistic and cultural differences and the increasing interest in connecting assessment results to universal and generic instructional responses speak to the de-personalization of these assessment practices. These assessment systems ignore what educators know about high quality, child-centered assessment and its role in literacy pedagogy and curriculum. The use and circulation of assessments for summative purposes operates at a nexus of childhood-literacy-theory that privileges Eurocentric, English dominant, colonizing, and narrow notions of education and literacy that deny other ways of being literate. Commercial interests, the testing and education publishing industry, and the commodification of testing, have contributed to the circulation of these assessments and accompanying neoliberal ideologies.

Formative Literacy Assessment: Global Patterns

Formative uses of literacy assessments also circulate across national borders and global spaces. These formative assessments focus on the reading and writing processes and strategies that children bring to texts rather than mastery of particular skills and abilities. Formative assessments involve carefully observing children as they engage in reading and writing activities. These activities might entail identifying or sorting letters, reading environmental print, or handling or reading books. Formative uses of literacy assessment involve watching children as they engage in activities designed to mimic everyday literacy activities and are premised on the belief that individual children may present a vast range of responses when asked to engage in a particular activity. Observation is key to discerning how individual children understand literacy tasks based on the strategies that they use to complete the task.

As Greenstein (2010) argues, formative uses of assessment are both student-focused and instructionally informative; they attend to individual children's responses to texts as means to identify and create instructional practices that will best serve children. For example, the implementation of Clay's (2019) *Observation Survey* might reveal that a child has difficulty hearing and recording sounds, learning particular letters or tracking print. Likewise, a running record or miscue analysis might suggest that a child over-relies on phonic cues and neglects meaning or syntactic information. Each of these insights would lead to different instructional practices designed to address particular challenges. Informal formative assessments can be built into instructional programs supporting rather than disrupting ongoing instruction.

The purpose of formative assessments is not to evaluate or calibrate children in relation to narrow and pre-determined models of expected growth. Instead, these assessments seek to discern what children understand about literacy and the degree to which they are able to use that knowledge to engage with text. Formative uses of literacy assessment that are used in a variety of contexts and systems around the world include: *The Observation Survey of Early Literacy Achievement* (Clay, 2019), running records (Clay, 2013), and miscue analysis of reading (Goodman et al., 2005). These assessments circulate globally and are used by expert literacy teachers to support young readers and writers. Here we focus on these assessments and their capacity to provide formative data about children's literacy learning across global contexts.

The Observation Survey

The *Observation Survey of Early Literacy Achievement* (Clay, 2019) originated in New Zealand as the assessment tool used in the Reading Recovery intervention. Originally based on Clay's dissertation study and first published as a comprehensive set of observational tasks in 1993, the most recent version in the result of several instantiations. *The Observation Survey* strives to engage teachers in close observation of children as they engage in literacy tasks that reflect the types of literacy tasks that are often familiar to young children. In short, the tasks involve identifying words and letters, observing children as they engage with books and written language, writing words and sentences, and reading simple texts.

While the tasks are standardized in the sense that there are guidelines for both administration and scoring, they are designed to reflect the unique capabilities of individual children. For example, children might identify letters by letter name, sound, or a word that begins with that letter. Children earn points for words written correctly in the language used by the task administrator or for words written in other languages. Likewise, teachers notice when children write sentences using letter/sound correspondences that reflect their knowledge of other languages. These assessments, because they strive to observe children as they engage in naturalistic literacy activities, involve observing what children can do and how they do it. Most of these tasks are either untimed or entail generous time limits. Thus, they honor the strategies and knowledge that children bring to literacy tasks rather than expecting particular responses to specific prompts, which are scored in relation to narrow and specific criteria.

The *Observation Survey* (2019) has circulated extensively through the English-speaking world, initially through the proliferation of Reading Recovery, and then through networks of literacy educators who have found the assessment tasks useful. The *Observation Survey* (2019) has been "re-constructed" for children who are learning to read in Spanish (Escamilla, Andrade, Basurto, & Ruiz, 1996), French and several other European languages, as well as Walpiri (an Australian Indigenous language), Hebrew, and Braille (Rodriguez, Hobsbaum, & Bourque, 2003). The authors of the Spanish version clearly state that this is not a direct translation of the survey tasks. Instead, the Spanish version is a "conceptual re-creation of the work of Clay from English to Spanish, that further considers how children who come into contact with two languages use those two languages to make sense of their world" (Escamilla et al.,1996, p. 2). Thus, re-construction of the Observation Survey was created with careful attention to the nuances of Spanish and English. Aspects of spoken and written languages are recognized as affecting what children do, how the assessments are administered, and how they are scored.

Running Records

While running records of children's reading come in diverse forms, one version originated as part of the *Observation Survey of Early Literacy Achievement* (Clay, 1993). Running records have assumed a life of their own, with an ever-increasing diversity of formats, processes, and administration and scoring techniques. Running Records are now used across and beyond English–speaking nations (i.e., Australia, Canada, France, Israel, United Kingdom, USA, New Zealand) within and beyond the Reading Recovery program. For example, running records are the primary informal reading assessment task used in the United States of America (see Fountas and Pinnell, 2006; Pinnell & Fountas, 1996). As the *Observation Survey* continues to be re-constructed into more and more languages, guidelines for conducting the Clay version of running records are also becoming available in multiple languages. For example, original children's books have been written by native Spanish-speaking authors to provide running record materials

in Spanish (Fountas & Pinnell, 2011). While running records attend to accurate reading, when properly used these assessments are "not just about right or wrong words" (p. 55). As Clay (2013) explains:

A Running Record needs to capture all the behavior that helps us interpret what the child is probably doing. Everything the child said and did tells us something; what his hands and eyes were doing, the comments he made, when he repeated a line of text and so on. The aim is this: after a Running Record the teacher should be able to 'hear the reading again' when reviewing the record. (p. 55)

Unfortunately, in practice the use of Running Records does not always entail this level of analysis, nuance, and thoughtfulness. Running Records, when administered and analyzed correctly should provide a window into the readers' processing of text, not a simple count of words read correctly. When used thoughtfully, running records have helped teachers around the world to attend to what readers bring to texts and how they negotiate the challenges they encounter in texts. Running Records are designed as a tool for attending to the constructive, intentional, and emerging abilities that children bring to texts, thus highlighting a vision of childhood literacy and literacy learning as creative, non-linear, and unique to individual children.

Miscue Analysis

For over four decades, Ken and Yetta Goodman have developed and refined miscue analysis as a window into reading processes (Y. Goodman, Watson & Burke, 2005). Originating in the United States during the 1960s, miscue analysis has been used to understand the reading process of students in grades K-12. The assessment technique has been used around the world and translated into several languages including English, Spanish, French, Greek, Japanese, and Chinese Mandarin. Miscue analysis has also been used with children with deafness and other auditory limits.

In 2005, Yetta Goodman and her colleagues (2005) revised reading miscue analysis techniques to create a set of teacher-friendly assessment practices. In addition, *Retrospective Miscue Analysis (RMA),* has been developed to use findings from miscue analysis to inform instruction from the assessments that not only to help readers develop strong reading processes but also to help children who have experienced difficulties with reading revalue themselves as capable readers (Goodman & Marek, 1996).

Miscue analysis documents what children do as they read. It allows for teachers to attend to children's use of various informational systems including linguistic systems (i.e., semantic, syntactic, graphophonic) and the pragmatic systems (i.e., situation, background knowledge, culture). In short, teachers compare the child's reading of a text (i.e., repetitions, inserted words, substitutions, corrections) to the actual text in order to understand what children do well as readers and what types of instruction might ensue. The essential question that teachers and children ask themselves is "Did what you read make sense?" The focus is on meaning construction rather than code breaking. Like running records, miscue analysis procedures allow children to bring a vast set of reading strengths and abilities to text.

Because the focus is on meaning construction, dialectical differences and even changes in wording and syntax that do not disrupt meaning are not treated as problems. In short, there is more than one way to effectively read a text. Thus, like running records, miscue analysis is designed to attend to children's constructive, intentional, and emerging abilities with texts. Miscue analysis asks teachers to view childhood literacy and literacy learning as unique to individual children and as reflecting differences in the language practices and experiences that children bring to text.

Conclusions about the Circulation of Formative Assessments

The Observation Survey of Early Literacy Achievement (Clay, 2013), running records (Clay, 2013), and miscue analysis (Y. Goodman et al., 2005) are all student-focused assessments that have had global reach and are used formatively to inform literacy instruction. These assessments highlight what children know and provide ways for teachers to learn from the errors that children make while engaging with text (Clay, 1975). When used formatively, these assessments allow for a wide range of responses including responses that reflect children's cultural and linguistic backgrounds. Commenting on evaluations of *No Child Left Behind*, Darling-Hammond (2007) worried that summative and high-stakes assessments increased the "likelihood that the most vulnerable students would be severely victimized by a system not organized to support their learning" (p. 6). Paul (2004) maintained that *No Child Left Behind* established accountability through testing that placed "black and Latino students in imminent educational danger" (p. 696) and had lifelong consequences in regard to "the workforce, the prison industrial complex, and higher education" (p. 649). As Cochran-Smith and Lytle (2006) argue, "teachers' actions are infused with complex and multilayered understandings of learners, culture, class, gender, literacies, social issues, institutions, 'herstories' and histories, communities, materials, texts, and curricula" (p. 691). We suggest that standardized assessments that focus solely on outcomes and are designed from afar do not provide this same level of nuanced understanding of children and the strengths of the communities to which they belong.

The significance of student diversity has real effects on conceptualizations of childhood and learning. Drawing on an analysis of assessment across the European Union, Dolin, Black, Harlen, and Tiberhein (2017) argue that some views of children and learning place children at the center and provide children a role in the assessment process which further enhances learning. Afflerbach (2016) concurs; he espouses an "assessment credo" that demands that not only should assessments provide "information that is useful", but they should also "do no harm" (pp. 413-414). Arguing for the power of formative assessments, he cites the need for assessments to attend to student differences and develop teacher expertise, as well as foster students' self-efficacy and meta-awareness of themselves as readers.

Effective and culturally sensitive use of these assessments requires teacher expertise; and thus, an investment in teachers and faith that when provided with information and opportunities to learn and think that teachers can better serve all children. Unfortunately, neoliberal responses to international concerns about children's literacy learning have resulted in scripted programs, standardized curricula, and skills-based assessments that can be easily scored and are assumed to measure children's literacy abilities.

Internationally, the marketing of teacher-proof assessments and instructional materials have proven to be financially lucrative across international educational markets (Bennett, 2013). While both summative and formative assessments circulate globally, they circulate in different ways and with different effects. Assessments that are used summatively tend to have corporate or international for-profit roots (Darling-Hammond, 2007; Garan, 2005). Corporations, including Pearson and Kaplan, have found the international distribution of reading and literacy assessments to be a lucrative and previously untapped market. Creating a single set of assessments that can be roughly translated and distributed around the world entails huge profits particularly when considering the accompanying need for test preparation, scoring, data analysis, and teacher training, and the adjunct markets of parent materials and tutoring schemes.

In contrast, formative assessments require less corporate investment and less corporate profit. Formative assessments circulate through informal networks, require minimal materials, and are scored by teachers. Thus, while formative assessments do involve profits for authors and publishers, they have a

different relationship to the economy of publishing. Assessments that are used formatively generally involve a small number of professional texts and are often introduced to teachers introduced in literacy methods courses. Professors, and later literacy coaches in schools, support teachers in implementing and analyzing formative assessment data as teachers develop their abilities to use assessments to design instruction for a vast range of students. Thus, developing expertise around the use of formative assessments is a local and lifelong process that involves an extensive community of learners and investments in teacher expertise.

While scholars have made compelling arguments for the compatibility of formative and summative assessments (Dixon & Worrell, 2016; Dolin, Black, Harlen, & Tiberhein, 2017; Wixson, 2017), the current privileging of the summative uses of assessments internationally has created not only a robust financial market for these products but also affected the ways literacy learning and childhood are conceptualized and treated across nexus' of childhood, literacy, and theory. Unfortunately, the formative uses of assessments and the ways they support individual children are often invisible. Both teachers' records of informal assessments and resulting instructional changes remain within classrooms, and are rarely shared or publically recognized. Similarly, the gains made by individual children remain unrecognized or are dismissed as anecdotal. In recent years, the privileging of summative assessment practices has limited the resources and time dedicated to developing teacher expertise with formative assessments and the professional capacity of teachers to use formative assessment data.

CONCLUSION

In this chapter, we have argued that the international circulation of largescale summative assessments presents a childhood-literacy-theory nexus that not only manifests problematic views of childhood, but also privileges Eurocentric, English-centric, and narrow notions of learning. Specifically, large-scale literacy assessments that are used summatively are grounded in neoliberal and mechanical views of childhood and literacy that invite data–driven scripted and programmatic instruction, invoke and respond to supposedly universal conceptions of childhood, ignore the linguistic and literacy repertoires of multilingual children, often undermine and negate teachers' professional judgement, and are driven by the investment of capital rather than the interests of children. We connect international flows of assessment practices and policies to international missions to measure learning in order to compare learners and their teachers, and alert readers to the operation of largescale summative assessments as colonizing forces. Our goal is to raise awareness of the childhood-literacy-theory nexus that operates through the uses of assessments. Finally, we argue for the affordances of formative assessment including opportunities to discover what children know, individualize and inform instruction, honor cultural and linguistic diversity, and contribute to the development of teacher expertise.

At the same time, we also recognize limitations that accompany the use of formative assessments and admit that the widespread use of formative assessments requires investment in schools and teachers. However, we simultaneously consider the massive cost of summative assessment systems, both economically and in terms of how literacy is conceptualized and how children are treated and taught. In short, we support educational policies and practices that entail organizational and instructional systems that nurture teacher expertise and support student learning by providing valuable information that can significantly inform teaching.

REFERENCES

Apple, M. W. (2011). Democratic education in neoliberal and neoconservative times. *International Studies in Sociology of Education*, *21*(1), 21–31. doi:10.1080/09620214.2011.543850

Australian Curriculum, Assessment and Reporting Authority (ACARA). (n.d.). The tests. Retrieved from https://www.nap.edu.au/naplan/the-tests

Bartlett, L., & Vavrus, F. (2016). A vertical case study of global policy-making: Early grade literacy in Zambia. The Handbook of Global Education Policy, 554-572.

Bennett, T. (2013). *Teacher proof: Why research in education doesn't always mean what it claims, and what you can do about it*. New York: Routledge. doi:10.4324/9780203782293

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5.

Buckingham, D. (2017). Pirls of wisdom the limits of educational data. Retrieved from https://davidbuckingham.net/2017/12/15/pirls-of-wisdom-the-limits-of-educational-data/

Calkins, L. M. (1983). *Lessons from a child: On the teaching and learning of writing. Inc*. Portsmouth, NH: Heinemann.

Chall, J. S., & Jacobs, V. A. (1983). Writing and reading in the elementary grades: Developmental trends among low-SES children. *Language Arts*, *60*(5), 617–626.

Clark, M. (2017). *Reading the evidence: Synthetic phonics and literacy learning*. Birmingham, UK: Glendale Education.

Clay, M. M. (1913). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann.

Clay, M. M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.

Clay, M. M. (2019). *An observation survey of early literacy achievement* (4th ed.). Auckland, N.Z.: The Marie Clay Literacy Trust.

Cochran-Smith, M., & Lytle, S. (2006). Troubling images of teaching in no child left behind. *Harvard Educational Review*, *76*(4), 668–697. doi:10.17763/haer.76.4.56v8881368215714

Darling-Hammond, L. (2007). Evaluating" no child left behind. *Nation*, *284*(20), 11.

Dixon, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into Practice*, *55*(2), 153–159. doi:10.1080/00405841.2016.1148989

Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring relations between formative and summative assessment. In *Transforming assessment* (pp. 53–80). Cham: Springer. doi:10.1007/978-3-319-63248-3_3

Edmondson, J. (2001). Taking a broader look: Reading literacy education. *The Reading Teacher*, *54*(6), 620.

Escamilla, K., Andrade, A. M., Basurto, A., & Ruiz, O. (1996). *Instrumento de Observacion de los Logros de la Lecto-Escritura Inicial*. Portsmouth, NH: Heinemann.

Flewitt, R., & Roberts-Holmes, G. (2015). Regulatory gaze and 'non-sense' phonics testing in early literacy. Negotiating Spaces for Literacy Learning Multimodality and Governmentality, In M. Hamilton, R. Heydon, K. Hibbert, & R. Stooke (Eds.), Negotiating spaces for literacy learning: Multimodality and governmentality (pp. 95-114). London: Bloomesbury.

Fontaine, C., & Luttrell, W. (2015). Re-centering the role of care in young people's multimodal literacies: A collaborative seeing approach. In M. Hamilton, R. Heydon, K. Hibbert, & R. Stooke (Eds.), *Negotiating spaces for literacy learning: Multimodality and governmentality* (pp. 43–56). London: Bloomesbury.

Foorman, B. R., & Connor, C. M. (2011). Primary grade reading. In Handbook of reading research (Vol. 4, pp. 136-156). Routledge.

Fountas, I. C., & Pinnell, G. S. (2006). *Teaching for comprehending and fluency K-8: Thinking, talking, and writing about reading*. Portsmouth, NH: Heinemann.

Fountas, I. C., & Pinnell, G. S. (2011). *Sistema de evaluación de la lectura*. Portsmouth, NH: Heinemann.

Gandini, L. (1993). Fundamentals of the Reggio Emilia approach to early childhood education. *Young Children*, *49*(1), 4–8.

Garan, E. M. (2005). Murder your darlings: A scientific response to the voice of evidence in reading research. *Phi Delta Kappan*, *86*(6), 438–443. doi:10.1177/003172170508600606

Goodman, K. (1996). *On reading*. Portsmouth, NH: Heinemann.

Goodman, Y. M. (1978). Kid Watching: An Alternative to Testing. *National Elementary Principal*, *57*(4), 41–45.

Greenstein, L. (2010). *What teachers really need to know about formative assessment*. Alexandria, VA: Association for Supervision & Curriculum Development.

Hammill, D. D., & Swanson, H. L. (2006). The National Reading Panel's meta-analysis of phonics instruction: Another point of view. *The Elementary School Journal*, *107*(1), 17–26. doi:10.1086/509524

Hibbert, K. M., Heydon, R. M., & Rich, S. J. (2008). Beacons of light, rays, or sun catchers? A case study of the positioning of literacy teachers and their knowledge in neoliberal times. *Teaching and Teacher Education*, *24*(2), 303–315. doi:10.1016/j.tate.2007.01.014

Hoffman, J. V. (2012). Standpoints: Why EGRA—a clone of DIBELS—will fail to improve literacy in Africa. *Research in the Teaching of English*, 340–357.

Hogan, A. (2016, January 11). NAPLAN and edu-business: The commercialization of schooling in Australia. *EduResearch Matters*.

Howie, S. J., Combrinck, C., Tshele, M., Roux, K., McLeod Palane, N., & Mokoena, G. M. (2017). *PIRLS 2016 Progress in International Reading Literacy Study 2016 Grade 5 Benchmark Participation: South African Children's Reading Literacy Achievement*. Pretoria: Centre for Evaluation and Assessment.

Kaminski, R. A., & Good, R. H. III. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, *25*(2), 215–227.

Klingner, J. K., & Harry, B. (2006). The special education referral and decision-making process for English Language Learners: Child study team meetings and staffings. *Teachers College Record*, *108*, 2247–2281. doi:10.1111/j.1467-9620.2006.00781.x

Land, S. (2015). *Reading isiZulu: reading processes in an agglutinative language with a transparent orthography.* Unpublished PhD Thesis, University of KwaZulu Natal, Pietermartizburg.

Montessori, M. (2013). *The Montessori Method*. New Brunswick, USA: Transaction publishers.

Mullis, I. V., Martin, M. O., Foy, P., & Drucker, K. T. (2012). PIRLS 2011 International Results in Reading. The Netherlands: International Association for the Evaluation of Educational Achievement.

Murphy, S. (2015). Beyond governmentality: The responsible exercise of freedom in pursuit of literacy assessment. In M. Hamilton, R. Heydon, K. Hibbert, & R. Stooke (Eds.), *Negotiating spaces for literacy learning: Multimodality and governmentality* (pp. 25–41). London: Bloomesbury.

National Reading Panel. (2000). *National reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Washington, DC: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development.

Noddings, N. (2005). What does it mean to educate the whole child? *Educational Leadership*, *63*(1), 8–11.

Pearson. (n.d.). Five facts about NAPLAN and Pearson. Retrieved from https://pearson.com.au/insights-and-news/pearson-news/five-facts-about-naplan-and-pearson/

Pinnell, G. S., & Fountas, I. (1996). *Guided reading: Good first teaching for all children*. Portsmouth, NH: Heinemann.

Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. New York: Vintage.

Read, C. (1986/2018). *Children's creative spelling*. New York: Routledge.

Rodriguez, Y., Hobsbaum, A., & Bourque, G. (2003). The reconstruction of the Observation Survey in other languages. *Journal of Reading Recovery*, 38-40.

Rogers, R., Mosley, M., & Folkes, A. (2009). Standing up to neoliberalism through critical literacy education. *Language Arts*, *87*(2), 127.

Shearer, A., & Klenowski, V. (2019). Assessment as discursive practice in the early years. In A. Woods & B. Exley (Eds.), *Literacies in early childhood: Foundations for equity and quality* (pp. 273–291). Australia: OUP.

Stevens, L. P. (2003). Reading First: A critical policy analysis. *The Reading Teacher*, *56*(7), 662–668.

Teale, W. H., & Sulzby, E. (1986). *Emergent Literacy: Writing and Reading. Writing Research: Multi-disciplinary Inquiries into the Nature of Writing Series*. Ablex Publishing Corporation.

Office of the Deputy Secretary, U.S. Department of Education. (2004). No child left behind: A toolkit for teachers. Retrieved from https://www2.ed.gov/teachers/nclbguide/nclb-teachers-toolkit.pdf

Van Rooy, B., & Pretorius, E. (2013). Is reading in an agglutinating language different from an analytic language? An analysis of isiZulu and English reading based on eye movements. *Southern African Linguistics and Applied Language Studies*, *31*(3), 281–297. doi:10.2989/16073614.2013.837603

Wixson, K. K. (2017). An interactive view of reading comprehension: Implications for assessment. *Language, Speech, and Hearing Services in Schools*, *48*(2), 77–83. doi:10.1044/2017_LSHSS-16-0030 PMID:28395296

Woods, A., & Comber, B. (2019). Bringing together what we know about literacy and equity: A discussion about what's important. In A. Woods & B. Exley (Eds.), *Literacies in early childhood: Foundations for equity and quality* (pp. 2–15). Australia: OUP.

Wyse, D., & Goswami, U. (2008). Synthetic phonics and the teaching of reading. *British Educational Research Journal*, *34*(6), 691–710. doi:10.1080/01411920802268912

Wyse, D., & Styles, M. (2007). Synthetic phonics and the teaching of reading: The debate surrounding England's 'Rose Report'. *Literacy*, *41*(1), 35–42. doi:10.1111/j.1467-9345.2007.00455.x

Yatvin, J., Weaver, C., & Garan, E. (2003). Reading first: Cautions and recommendations. *Language Arts*, *81*(1), 28–33.

ENDNOTE

[1] In South Africa grade 5 children took the PIRLS test.