

Investigating the comprehension iceberg: Developing empirical benchmarks for early-grade reading in agglutinating African languages

**Authors:**

Nicholas Spaull¹ 
 Elizabeth Pretorius² 
 Nompumelelo
 Mohohlwane³ 

Affiliations:

¹ReSEP, Department of Economics, Stellenbosch University, Stellenbosch, South Africa

²Department of Linguistics and Modern Languages, University of South Africa, Pretoria, South Africa

³Department of Basic Education, Research Coordination, Monitoring and Evaluation, Pretoria, South Africa

Corresponding author:

Nicholas Spaull,
 nicholasspaull@gmail.com

Dates:

Received: 04 May 2019
 Accepted: 02 Nov. 2019
 Published: 18 Feb. 2020

How to cite this article:

Spaull, N., Pretorius, E. & Mohohlwane, N., 2020, 'Investigating the comprehension iceberg: Developing empirical benchmarks for early-grade reading in agglutinating African languages', *South African Journal of Childhood Education* 10(1), a773. <https://doi.org/10.4102/sajce.v10i1.773>

Copyright:

© 2020. The Authors.
 Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Read online:

Scan this QR code with your smart phone or mobile device to read online.

Background: Reading development in agglutinating African languages is a relatively under-researched area. While numerous studies highlight the low comprehension levels among learners reading in African languages in South Africa, little has been done to probe beneath this 'comprehension iceberg' in terms of decoding skills.

Aim: As a tentative step towards benchmarking in African languages, we analyse the sub-components of reading across three languages (Northern Sotho, Xitsonga and isiZulu), to better understand the nature of alphabetic knowledge, word reading and fluency in these languages, how these relate to one another, and how accuracy and speed relate to comprehension.

Setting: Data was obtained from 785 Grade 3 learners across three African languages in three provinces in South Africa.

Methods: The early grade reading assessment (EGRA) framework was adapted to the written features of the three languages to assess letter-sounds, single-word reading, non-word reading, oral reading fluency (ORF) and oral comprehension.

Results: We present results on fluency, accuracy and comprehension and their interrelationships in these morphologically rich languages. While differences emerged between the conjunctive and disjunctive orthographies, strong relations occurred across the languages between letter-sound knowledge and word reading, word reading and oral reading fluency, and ORF and reading comprehension. Results suggest minimum thresholds of accuracy and ORF in each language, below which it is virtually impossible to read for meaning.

Conclusion: There is a strong need for language-specific norms and benchmarks for African languages. Preliminary minimum decoding thresholds for comprehension found in these three languages serve as a move in that direction.

Keywords: early literacy; EGRA; letter sound knowledge; word reading; oral reading fluency; reading comprehension; African languages; reading benchmarks; South Africa.

Introduction

Given the important role of reading in scholastic performance, it is important to ensure that children are launched on successful reading trajectories from the start of schooling. The Progress in International Reading Literacy Study (PIRLS) is an international reading comprehension assessment conducted at Grade 4 levels in which South Africa has participated for over a number of years (2006, 2011 and 2016). The PIRLS results show that Grade 4 children in South Africa perform very poorly in reading comprehension, even when reading in their African home language (Howie et al. 2006, 2012, 2017; Spaull & Pretorius 2019). The latest round (2016) showed that 78% of South African Grade 4 learners have not learnt to read for meaning in any language by Grade 4 (Howie et al. 2017), suggesting unsuccessful reading trajectories. Yet, this does not tell us much about which components of the reading processes children are struggling with. In this article, we analyse data from one-on-one assessments of Grade 3 learners across three languages (Northern Sotho, Xitsonga and isiZulu) with the aim to better understand the levels and distributions of these underlying reading processes.

There are several factors that contribute to children's ability to read for meaning. These include the richness of children's vocabulary and grammatical knowledge of the language in which the text is written, the technical ability to decode the written text, home factors and formal school practices that enculturate children into reading and making meaning of written texts, as well as children's

familiarity with different written genres of text and use of strategies of how to 'read them'. Thus, while the PIRLS results provide information about reading comprehension, and clearly signal challenges within the education system regarding comprehension, they do not tell us why South African children struggle with reading comprehension. The very high number of children who could not read for meaning at all (78%) also raises questions *inter alia* about the development of early decoding skills, and how these enable comprehension, particularly in the African languages.

Decades of research into reading in English has provided education stakeholders with an evidence-based framework for profiling what successful reading in English looks like (Adams 1990; National Reading Panel 2000). For example, by the end of Grade 3 children at the 50th percentile in the United States can, on average, read 107 words correct per minute (wcpm) in English (Hasbrouck & Tindal 2006), while children reading slower than 40 wcpm at the end of Grade 1 are considered to be at reading risk. This information in and of itself cannot tell us anything directly about reading comprehension, but given the robust correlations typically found between reading rate and reading comprehension (Cummings & Petcher 2016), it enables teachers to deduce with some measure of confidence that if readers with English as their home language are only reading at 60 wcpm in Grade 3, they are likely to struggle to read with meaning *on their own*¹ (Hasbrouck & Tindal 2006). Notwithstanding the importance of this contribution to our general understanding of reading in alphabetic languages, identifying what is generic and what is language-specific in early reading development calls for a research base that includes alphabetic languages that are typologically different and have different orthographic systems. The African languages spoken in South Africa are agglutinating, syllabic languages with a transparent orthography, as opposed to English being a partially analytic, stress-timed language with an opaque orthography. What would an average Grade 3 or an at-risk Grade 1 reader in an African language look like? At present, we cannot say for sure because relatively little reading research has been conducted in these languages (Pretorius 2018). Currently, anecdotal experience, intuitions and linguistic hunches tend to underlie educational judgements about how young African language readers are faring. In many cases, teachers are poorly trained in how to teach reading (Taylor & Taylor 2013) and do little reading themselves (Pretorius & Knoetze 2012).

South Africa has prioritised the large-scale measurement and monitoring of reading comprehension outcomes across the country². While there are several nuances in the successive

1. Note that learners may follow a text and be able to answer various comprehension questions orally if the text is read to them and the questions are mediated by a teacher. That, however, falls within the domain of listening comprehension: reading comprehension requires learners to read and answer questions on a text on their own.

2. These include PIRLS, the Southern and East African Consortium for Monitoring Education Quality (SEACMEQ), Annual National Assessments (ANA) and the National Integrated Assessment Framework (NIAF) referred to as the National Assessment Programme (NAP). These are undertaken nationwide by the Department of Basic Education.

results of the large-scale comprehension assessments undertaken in South Africa, what is lacking is not accurate information on reading outcomes but accurate information on what is less visible beneath the comprehension iceberg. As De Vos, Van der Merwe and Van der Mescht (2014:168) point out, very little has been done on the 'cognitive-linguistic processes involved in reading in African languages'. We do not yet know what successful early reading trajectories look like in African languages, and how they are similar or different across the different African languages. The metaphor of the comprehension iceberg refers to what is still unknown and invisible beneath reading comprehension results in the South African context. A strong empirical base is needed to gain insight into early reading development in African languages and make sound judgements about ways to reduce the literacy inequalities within the education system.

Given the relative paucity of research on decoding in African languages, this article uses Grade 3 reading data from three African languages in South Africa to examine the nature of alphabetic knowledge, word reading and oral reading fluency (ORF) in these languages (looking at means and dispersion within the cohort), how these relate to one another and how accuracy and speed in different decoding components relate to one another and to reading comprehension. These findings are then used to identify minimal decoding thresholds, below which reading comprehension is difficult to achieve. This is a first, tentative step towards suggestions for benchmarking in African languages.

Focusing attention on these foundational reading skills does not imply that decoding forms the bulk of the comprehension iceberg. Indeed, Snow and Kim (2007) state that decoding skills comprise a small problem space in relation to the large problem space of vocabulary and comprehension. However, if skills in the small problem space are not well developed, then reading comprehension is severely compromised. As Adams (1994:838) has long argued, if decoding skills such as word recognition do not operate properly, 'nothing else in the system can either'.

Before turning to the research, we first discuss some features of early reading in alphabetic languages, briefly outline ways in which African languages differ from English and the implications this may have for reading, and then look at the role of alphabetic knowledge, word reading and ORF in early reading development.

Early reading development in alphabetic languages

The first 3 years of schooling are typically dedicated to laying a sound foundation for the development of numeracy and literacy skills on which all subsequent schooling depends. By the end of Grade 3, readers are expected to read accurately – on their own – at a steady rate or speed (appropriate to their grade level), with comprehension and with enjoyment. Comprehension is the sine qua non of reading: we read to

comprehend information in written text. The aim of reading instruction is for children to understand how the written code represents spoken language, and to be able to decipher the code in any text (i.e. decode) accurately and rapidly so that meaning can be constructed from the text when children read on their own.

Accuracy in decoding supports comprehension. The ability to identify letter-sounds accurately and use this knowledge to read words accurately reduces comprehension complications (Adams 1994; Spear-Swerling 2006). For example, it is important to distinguish *three* from *tree* in English, or *abafundi* [learners] from *bafunda* [they read] in isiZulu. Speed also matters in text processing, and hence in reading. A difference of a few milliseconds can signal difficulty or success in text processing. While speed does not translate directly to reading comprehension (children may read quickly without understanding what they are reading), processing speed tends to be strongly associated with word reading and reading comprehension (Cummings & Petscher 2016; Fuchs et al. 2001; Siedenberg 2017; Wolf & Katzir-Cohen 2001). As children become more accurate in their decoding, they process words more quickly as their eyes move across the text. The more the effort expended on processing the alphabetic code and words, the less is the attentional capacity for comprehension. Finally, affect and motivation are linked to reading; children who enjoy reading are more likely to engage in reading, do more reading and hence become better readers (Guthrie et al. 2007).

Research into the acquisition of literacy has shown that individual differences between learners in accuracy, speed and comprehension can emerge early, and if weaknesses in these areas are overlooked and not remediated, reading problems will persist throughout schooling (Spear-Swerling 2006). Such differences also create spill-over effects. If some children find reading effortful and frustrating, they will not perceive it as meaningful or pleasurable, and therefore would be less inclined to actively engage in it. It is difficult to interpret reading differences between children unless one knows what typical or (un)successful reading trajectories look like at different grade levels. Thal et al. (1997:241) argue that ‘if there are no clear criteria for identifying what is “normal”, then it is especially difficult to be certain that a child is delayed or precocious’. For example, if a learner reads at 25 wcpm, is he or she a good, average or struggling reader? If we are further told that the learner is in Grade 2 and is reading at 25 wcpm in isiZulu, can we now tell if he or she is a good, average or struggling reader? And if a Grade 2 learner reads at 25 wcpm in Northern Sotho, does this change our assessment of how well he or she reads? Given the paucity of research on early reading trajectories in different African languages, it is very difficult to know whether a Grade 2 learner reading at 25 wcpm is a delayed, average or precocious reader, and whether the African language in which he or she reads makes a difference to this assessment. These are the as-yet unobserved and unexamined issues underlying the reading comprehension iceberg in the South African context that this article addresses. In particular, it examines whether the relationship between

accuracy, speed and reading comprehension plays out in different ways in different African languages.

Typological and orthographic features of agglutinating African languages

This section highlights some features that distinguish agglutinating African languages and their orthographies from English, and identifies in what ways these features might impact early reading development.

Agglutinating languages: Morphological complexity

In linguistic typology, a distinction is made between isolating, inflectional and agglutinating languages, depending on the extent to which morphemes are added to stems or roots by way of prefixes, infixes and suffixes to mark grammatical meanings, such as gender, number, tense, aspect and locus. The differences between these categories are not absolute; rather, morphosyntactic features can be viewed as being on a continuum of lesser or greater complexity, with isolating languages having less and agglutinating languages having greater morphological density. English has features of an isolating language while languages such as Latin, Greek, Spanish and German are regarded as inflectional languages. The nine African languages spoken in South Africa belong to the family of Southern African Bantu languages. They are all agglutinating languages with a rich and complex morphology whereby prefixes, infixes and suffixes are added to noun and verb roots. The verbal elements in a sentence are especially complex, marking subject and object noun class, but they do so in slightly different ways, depending on the different language family clusters. The nine South African languages are divided into the Nguni (isiZulu, isiXhosa, Siswati and isiNdebele) and Sotho (Northern Sotho, Southern Sotho and Setswana) subfamilies, and two smaller subfamilies (Tshivenda and Xitsonga, related to languages in Zimbabwe and Mozambique). The reading data presented in this article were collected for isiZulu ($n = 514$), Northern Sotho, also called Sepedi ($n = 143$) and Xitsonga ($n = 128$) Grade 3 readers, and thus reflect the three main linguistic subgroups, as highlighted in Figure 1.

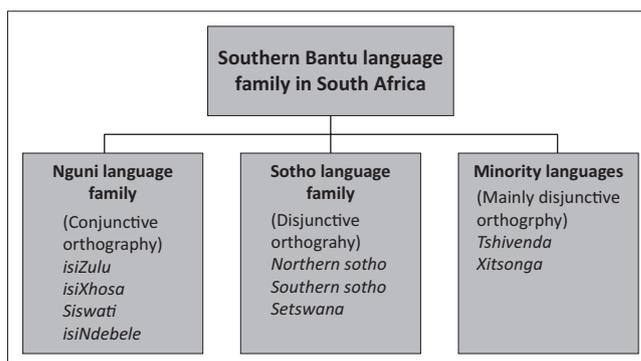


FIGURE 1: The Southern Bantu language families in South Africa.

Other agglutinating languages include Finnish, Turkish and Basque. Morphological complexity is a distinctive feature of all these languages, and a single orthographic word with a stem and morphemes stacked onto it can represent a whole sentence. In Turkish and Finnish, the morphemes tend to stack up as a series of suffixes attached to the root (Miller, Guldenoglu & Kargin 2019; Silven, Poskiparta & Niemi 2004), while in African languages the morphemes stack up before and after the root, as prefixes, infixes and suffixes. For example, the word *Andizithandi* in isiXhosa ('I don't like them', i.e. cakes) has the root *-thand(a)-* 'like' with the morphemes *a-ndi-zi* and *-i* attached: the prefix *a* and the suffix *i* signal the negative; *ndi* denotes first-person singular subject and *zi* refers to 'them' (cakes) that belong to the *isi-* or *izi-* noun class denoting, *inter alia*, foodstuffs. The agglutinative nature of these languages has implications for early reading development, as explained further.

Writing systems: Transparency and orthographic boundaries

Phonology refers to the sound system of a language, while *orthography* refers to the way in which oral or spoken language is represented in written form. In languages with an alphabetic orthography, spoken language is represented in written form at the phonemic level. In other words, written symbols (i.e. the various letters of the alphabet) represent the distinctive sound units of a language (i.e. phonemes). While there are 26 basic letters of the alphabet, languages use these letters in different configurations to map them to their particular phonology. For example, the phoneme /tʃ/ occurs in various languages: in English the sound is represented by the letters 'ch' or 'tch' as in *church* or *watch*, while the same sound is represented by the letters 'tsh' in Northern Sotho (*tshela* [cross over]), Xitsonga (*tshama* [sit]) and Zulu (*utshani* [grass]).

Orthography is transparent in all nine African languages – this means that letters represent specific sounds in a one-to-one mapping relationship. This is unlike English with its opaque orthography, where one letter can represent different sounds (the vowel *a* is sounded differently in *car*, *call*, *cane*, *alone*), or where the same sound can be represented by different letters (/f/ can be written as *f*, *ph*, or *-gh* in *frog*, *phone* and *cough*). Seidenberg (2017:136) points out that languages with complex morphological systems all have

transparent orthographies; an inconsistent orthography in agglutinating languages would make reading 'intolerable'.

Although the orthography in African languages is transparent, a distinction is made between conjunctive and disjunctive orthographies. This distinction coincides with language family groupings; the Nguni languages having a conjunctive orthography and the Sotho languages have a disjunctive orthography.

Morphophonological features specific to the different African languages (e.g. vowel elision in the Nguni languages) have influenced to some extent the development of different transparent orthographies for these languages. For example, in the conjunctive orthography of the Nguni languages, nominal and verbal elements in a sentence tend to be written together as single orthographic 'words'. In contrast, the Sotho languages evolved a disjunctive orthography, where some of the verbal elements in a sentence (e.g. noun class markers and suffixes) are written separately. For example, 'They used to read it' is written conjunctively as a single orthographic word *Babeyifunda* in isiZulu, while it is written disjunctively as five separate words *Ba be ba e bala* in Northern Sotho. Xitsonga orthography is somewhere in between, having elements of both conjunctive and disjunctive orthography, as shown in Table 1. The conjunctive or disjunctive distinction has implications for early reading, measurement and benchmarks.

Conjunctive orthography gives rise to long word units that create 'dense' texts; conversely, disjunctive orthography results in much shorter word units (often single syllables comprising vowel (V) or consonant vowel (CV). Because of its conjunctive orthography, there are typically few free morphemes in a Nguni language sentence – bound morphemes by way of prefixes, infixes and suffixes are added to noun and verb stems. Single-syllable words are practically non-existent (they are mainly exclamations) and two-syllable words are not common in the conjunctive orthography. Because of the noun class prefix attached to a noun stem, nouns typically contain three or more syllables. In terms of text length, equivalent texts translated into the conjunctive Nguni texts will yield short texts with long words, while the same text in a disjunctive Sotho language will yield longer texts with many short words. To illustrate these orthographic differences, examples taken from the first three sentences in a Grade 3 reader, in isiZulu, Northern Sotho and Xitsonga, respectively, are given in Table 1.

TABLE 1a: Words per sentence in conjunctive or disjunctive orthographies.

Language	Text
Northern Sotho	<i>Ka le lengwe la matšatši mosepedi yo a bego a na le tlala. O fihlile motseng wo mongwe a kgopela dijo. Go be go se na yo a bego a na le dijo</i>
Xitsonga	<i>Siku rin'wana mufambi loyi a ri na ndlala. U fikile emugangeni. A kombela swakudya, kambe a ku nga ri na loyi</i>
isiZulu	<i>Kunesihambi esasilambile kakhulu. Sahamba sicela ukudla emizini yabantu. Abantu abengenakho ukudla</i>
Gloss	There was a stranger who was very hungry. He came to a village and asked for food. Nobody had any food

TABLE 1b: Words per sentence in conjunctive or disjunctive orthographies.

Language	Words in sentence 1	Words in sentence 2	Words in sentence 3	Total words	Words per sentence	Letters per word	Total single-syllable words: V/CV
Northern Sotho	13	8	12	33	11	3.2	21
Tsonga	8	3	10	21	7	4	9
Zulu	3	5	3	11	3.6	8	0

As can be seen, similar sentences yield texts with different profiles. While the first sentence contains 13 words in Northern Sotho and 8 in Xitsonga, there are only 3 long words in the isiZulu text. In isiZulu, the words are longer and sentences are shorter on average, compared to Northern Sotho and Xitsonga. The agglutinating nature of African languages, their complex consonants and the conjunctive or disjunctive orthographies may have important implications for reading development in these languages.

Foundational reading skills

To optimise reading instruction and to look out for those who fall behind their grade peers, it is important to understand the dynamics of how the different components of decoding and comprehension interact and mesh, and where and why reading fallout happens. Skills that are key to learning to read the alphabetic code are foregrounded in the initial stages of learning to read and may predict early reading skill in grades 1 or 2. When mastery is achieved, these skills become automatised and recede to the background, while qualitatively different processes and skills (e.g. inferencing, perspective taking and comprehension monitoring) become foregrounded and push reading development to another level (Adams 1990; Castles, Rastle & Nation 2018; Spear-Swerling 2006). The ways in which these components interact may be sensitive to specific linguistic and orthographic constraints associated with different languages that share the same alphabetic code.

Alphabetic knowledge

Alphabetic knowledge refers to knowledge that written symbols stand for the phonemes of the spoken language. Inability to grasp this principle negatively affects the development of decoding (Nieto 2005). This is typically assessed as knowledge of letter–sound relationships. Letter–sound knowledge is a critical foundational skill of early literacy acquisition (e.g. Muter & Diethelm 2001) and becomes the main processing stage in word reading, where children use their letter–sound knowledge to sound out new words not previously encountered. Letter–sound knowledge is also related to phonological awareness, especially at the phonemic level, which has been found to be important in learning to read across alphabetic languages. When children learn letter–sound relationships, they develop an awareness of individual sounds within words (Ziegler & Goswami 2005).

Letter–sound knowledge is a critical foundational skill of early literacy acquisition (e.g. Muter & Diethelm 2001) and becomes the main processing stage in word reading, where children use their letter–sound knowledge to sound out new words not previously encountered. Blaiklock (2004) suggests that the role between phonological awareness and reading development is mediated by letter knowledge.

Because of its strong link with early reading instruction, alphabetic knowledge seems to have a narrow developmental

window (Ouelette & Haley 2013). Using measures of alphabetic knowledge with preschool children can lead to floor effects (Burgess & Lonigan 1998), while using it with older learners can produce ceiling effects³ (Wise et al. 2007). However, given the slow rate of reading development and the low literacy levels that are usually prevalent in developing countries, assessing alphabetic knowledge with older learners may help to distinguish readers from non-readers who have not yet grasped the relationship between print and sound.

Word and non-word reading

The most basic task of reading is being able to process the meaning of individual words from print, their relationship to other words in a sentence and to construct the overall meaning of the text in which the words and sentences occur. Although the ability to read words quickly and accurately is but one aspect of reading, successful text reading and comprehension rest on this ability. In alphabetic scripts, this is not possible without initial letter–sound knowledge (Adams 1994; Share 1995). However, to build fluency, children need to become aware of recurring letter patterns in their own language, based on morphological and orthographic information, incorporating smaller and larger word chunks until full word recognition is reached (Castles et al. 2018; Ehri 2005; Share 1995). After several encounters, words become known and familiar, readers recognise word chunks and so build up word-specific knowledge that helps to speed up and automatise the reading process so that attention is freed up for comprehension. Over the decades, a number of researchers have shown a strong association between speed and accuracy of word reading and reading comprehension (e.g. Adams 1990; Perfetti 2007; Stanovich 1986).

Assessing children's word reading ability is a good way to assess their decoding ability. Context-free word reading by way of word lists containing increasingly longer and more complex words is a significant predictor of reading (Jenkins et al. 2003). The use of non-words is also commonly used to assess decoding ability. Non-words meet the phonological criteria of a language but do not exist, for example, *brillig*, *slithy* and *toves* in English. Because these words lack meaning and readers have no orthographic representations of such words, non-words eliminate lexical processing and reveal a reader's phonological recoding ability. Research shows that real words are processed faster and more accurately than non-words. This seems to apply not only in opaque orthographies but also in transparent agglutinating languages such as Turkish (Miller, Kargin & Guldenoglu 2014; Miller et al. 2019). This aspect has not yet been explored in African reading research.

Because of the opaque orthography of English, developing rapid and accurate word reading skills takes longer time than in transparent orthographies like Greek, Welsh, German or Spanish. In these languages, letter–sound mapping occurs

3. Floor and ceiling effects are obtained when scores either cluster towards the bottom end of the distribution (floor), suggesting that a test is too difficult for the testees, or when scores cluster towards the top end of the distribution (ceiling), suggesting that a test is too easy (Richards & Schmidt 2010:61).

without much difficulty because of its regularity, and children can become efficient decoders within a year or so (Ellis & Hooper 2001; Aro & Wimmer 2003; Ziegler & Goswami 2005). This has also been found in agglutinating languages such as Turkish (Babayağit & Stainthorp 2007; Oney & Durgunoglu 1997). In their study of differences in reading long, inflected words in Basque (an agglutinating language) Acha, Laka and Perea (2010) found that while Grade 3 children relied mainly on letter-sound decoding, word identification was faster and more efficient with Grade 6 readers, who besides phonological decoding seemed also to rely on basic orthographic and inflectional patterns in the language as they became exposed to less frequent words during reading. Children learning to read in African languages thus have an advantage in reading a transparent orthography, but this advantage may be offset by a complex consonant sound system and ineffective classroom practices.

Oral reading fluency

Oral reading fluency reflects the speed, accuracy and naturalness that readers display when reading a text aloud, following the intonation and rhythm of spoken language. Oral reading fluency is seen as a general indicator of reading competence (Cummings & Petcher 2016; National Reading Panel 2000). Because intonation is more difficult and subjective to assess, speed and accuracy form the main focus of ORF assessment. Typically, readers are given a text to read within a minute, with errors subtracted from the total number of words read in a minute, giving a score of wcpm. To control for decoding without understanding, a short oral reading comprehension follows.

Research shows a strong association between ORF and reading comprehension (Fuchs et al. 2001; Spear-Swerling 2006) despite differences in socio-economic status and instructional programmes; it occurs in children without reading difficulties as well as those with learning disabilities (Wolf & Katzir-Cohen 2001). It has also been found in second-language reading (Jimerson et al. 2013), also in South Africa, the country of analysis here (Draper & Spaul 2015; Pretorius & Spaul 2016).

Because ORF scores can be affected by several factors such as text difficulty, practice effects, topic and genre familiarity, an ORF score should ideally be derived from the mean obtained from reading more than one text. However, the *relationship* between ORF and reading comprehension seems to still prevail, despite variations in these factors. For example, in their assessment of ORF in Kiswahili and English, Piper and Zuilkowski (2016) tested 4385 grades 1 and 2 learners and found strong correlations between ORF and reading comprehension, on both timed and untimed tasks. They also found that learners did not perform significantly better in ORF or comprehension levels when the assessment was untimed.

The greatest growth in ORF seems to occur in the early school years between grades 1 and 4. Oral reading fluency is useful

for measuring small increases in improvement, unlike many other standard measures of performance that can only detect large changes in the outcome (Blachowicz et al. 2006). Typically, from Grade 4 onwards, the effects of ORF start to level off (Fuchs et al. 2001; Spear-Swerling 2006). Once reading becomes relatively fast and accurate, other variables account for differences in reading comprehension, such as vocabulary knowledge, inferencing abilities and text, genre and background knowledge.

Oral reading fluency norms have been established for English readers, showing how children from different grades and with different percentiles typically perform. However, very little research has been conducted on ORF in African languages.

Research on early reading development in African languages

In approximately 70% of primary schools in South Africa, children complete the first 3 years of schooling in their African home language, with English taught as an additional language (Pretorius & Spaul 2016:1450). The situation then changes from Grade 4 onwards, with 90% of all learners now learning with English as the medium of instruction (the remaining 10% learn in Afrikaans), with African languages taught as a home language subject. As the majority of learners learn to read and write in an African language, one would expect much of the research on early reading in South Africa to focus on reading in the African languages, or on early bilingual reading in an African language and English.

However, there are currently not many studies on early reading in African languages and a rather uneven picture emerges from them as not all studies focus on the same factors, use the same measures or use similar measures in the same way (e.g. some studies use timed word reading measures and others do not). Research findings from the Nguni (isiZulu and isiXhosa) and Sotho (Northern Sotho and Setswana) languages are available, but often come from small-scale studies, as indicated in an annotated bibliography on research on reading in African languages (Pretorius 2018), and as yet no research seems to have been conducted on reading in Xitsonga.

Letter-sound knowledge: Because there are many complex consonant sounds in African languages, with many digraphs (*hl, ph, tj*), trigraphs (*tsh*) and also four-letter consonant sequences (*mpfhl, ntlh*), it is important that children learning to read in African languages master these consonants. Children learning to read in African languages need to be able to distinguish between the different letter shapes, their sounds and combinations in order to get on with the task of learning to read words that combine single consonants, digraphs and trigraphs. Surprisingly, however, only a few studies have included measures of alphabetic knowledge in their assessment of early reading skills in African languages. These include Setswana (Lekgoko & Winskel 2008; Taylor et al. 2017), Northern Sotho (Wilsenach 2015) and Kenyan

African languages (Kim & Piper 2019). In the South African studies, with the exception of the EGRS I study (Taylor et al. 2017), the sample sizes are relatively small, ranging from 36 to 108 learners. Although these studies have shown a relationship between letter–sound knowledge and early literacy in African languages, the relationship has not yet been examined closely.

Word reading and ORF: Results on word reading and ORF in both Nguni and Sotho languages can be gleaned from a few studies. The three studies looking at Nguni languages are those of Pretorius (2015: isiZulu Grade 4), Diemer (2015: isiXhosa Grade 3) and Rees (2016: isiXhosa Grade 3). All three studies found low ORF scores (19 wcpm), with Pretorius also finding low accuracy (53% of words read correctly) and Diemer low comprehension levels (23% on the assessment). The two authors analysing Northern Sotho that has a disjunctive orthography were Wilsenach (2013, 2015) and Makaure (2017) who both found higher rates of accuracy (67% – 79% on word reading) and slightly higher ORF scores (35 wcpm), compared to the findings in the Nguni studies. Performance on single-word and text-word reading was highly correlated in Northern Sotho ($r = 0.78$).

It is clear that while interest in early reading in African languages is emerging, there are still many issues that need to be further researched:

- There are surprisingly few studies that directly examine the role of alphabetic knowledge in early African language reading.
- There has been little research in African language reading on the relationship between reading words and non-words.
- Although English reading research shows strong correlations between word reading and ORF measures with comprehension, in early African language reading the relationship varies from mild to robust.
- Despite their transparency, conjunctive or disjunctive orthographies seem to affect early reading rates differentially. The reading rates from the Nguni studies are uniformly slow, while the reading rates from the Sotho languages are relatively faster. However, there is as yet no clear picture of the range of performance at different percentiles within the different languages.
- No studies to date have examined possible decoding thresholds in the different African languages, below which it is difficult for learners to make sense of a text when they read on their own.
- Many of the studies reviewed involve fairly small sample sizes from a small number of schools (never more than 4–5 schools); therefore, generalisation is constrained. A much larger and more varied empirical base is needed for theory building and benchmarking.

This article presents findings that address these issues and proposes, as a tentative first step in the direction of benchmarking, an approach that identifies possible decoding thresholds across the three different orthographic languages that enable reading comprehension.

Methodology

In February 2017, Grade 3 reading data were collected and analysed from 61 schools across three provinces in South Africa, representing both conjunctive and disjunctive transparent orthographies.

Background of the study

The data presented in this article draw on a larger study formally known as the ‘Leadership for Literacy’ project. The schools selected for the study are typical of those which serve the majority of learners and come from three South African provinces – Gauteng, KwaZulu-Natal and Limpopo. Of the 61 schools in the study, 56 are from the poorest 60% of schools in the country (quintile 1–3) which are no-fee schools, and 5 are from Quintile 4 where some charge relatively low fees (<R3000/year; €190/year). The aim of the sampling process was to ensure that there was the full range of performance across quintile 1–3 schools in these provinces. A matched-pair design was used where the (allegedly) highest performing schools, as reported by government officials and non-governmental organisations (NGOs), were matched with socio-economically-similar schools in the area (Wills 2017). In reality, all no-fee schools performed poorly. In total, there were 21 schools from Gauteng, 21 in KwaZulu-Natal and 19 in Limpopo.

Because the Grade 3 reading test was part of a larger study, where leadership issues were also being assessed (not reported in this article), time and budget constraints limited the amount of reading data that could be collected. Only part of one morning was set aside per school for reading assessment, so the test could not be too long, and only about 15 Grade 3 learners could be assessed in the allotted time per school.

Procedures

The data were collected between February and March 2017 in all three provinces. Ethical approval for the study was obtained from the university and the national and provincial education departments. Parents were sent letters of consent and learners gave either verbal or written consent. The reading tests were administered one-on-one by trained fieldworkers, with information captured electronically on tablets using an early-grade reading assessment (EGRA)-specific software (Tangerine). Each test was designed to be completed within 15 min. In all, 785 Grade 3 learners were assessed: 514 in isiZulu, 143 in Northern Sotho and 128 in Xitsonga. All fieldworkers were native speakers of the language they were assessing, held at least a bachelor’s degree or 3 year diploma and received 3-days of intensive training.

Grade 3 reading assessment

The Grade 3 learner assessment was an adapted form of the EGRA that already existed for these three African languages⁴.

4. RTI International, together with reading experts, developed the Early Grade Reading Assessment known as EGRA, funded by USAID, the World Bank and other donors.

Each home language assessment consisted of five subtests: (1) a timed letter–sound subtest specific to each language containing rows of letters that learners must sound aloud; (2) a timed word subtest, consisting of a list of words that learners must read out aloud; (3) a timed non-word subtest; (4) reading the title of the ORF passage story and (5) an ORF passage read aloud within 1 min. Following the ORF subtest, learners were asked oral comprehension questions based on the passage. Various opt-out rules were applied in the various subtests to protect learners who could not read at all.

In each of the assessed languages, the letter sound section had 110 items, with 10 letters per row. In addition to the standard EGRA test of lowercase and uppercase letters, this subtest was adapted to include, after the first row, an array of digraphs, trigraphs and quadgraphs that characterise the complex consonant sounds of the African languages, interspersed amongst the single letters. (The effect of these complex consonant sounds on letter–sound fluency is not examined here but is the topic of another paper.)

Across the three languages, for both the word and the non-word reading tasks, there were 60 words per task, starting with shorter words and ending with longer words (e.g. from *ikati* to *intothoviyane* in isiZulu; from *pula* to *kanagelokopana* in Northern Sotho; and from *teka* to *mpfampfarhuta* in Xitsonga). In order to keep the word reading tests comparable across the three African languages, no single-syllable function words, uncommon in the conjunctive orthographies but common in the disjunctive Sotho orthographies, were included in the Northern Sotho word lists (e.g. *a*, *na*, *go*, *le* etc., as shown in

Table 1 earlier). The words in all three language word reading tasks were nouns or infinitive forms of verbs, ranging in length from two to seven syllables.

The ORF passage was ‘Rock Soup’, a narrative text from a South African graded-reading series (Vula Bula). Although this was the same story, given the conjunctive or disjunctive features of the three languages, there were 120 words in the Northern Sotho passage, 105 in the Xitsonga passage and 67 in the isiZulu passage.

Ethical considerations

Ethical approval for this study was obtained from the University of Stellenbosch (ethical clearance number SU-HSD-003116).

Data results and analysis

Table 2 reports a range of descriptive statistics for each of the EGRA subtasks by language group, including the number of learners in the sample, the 10th, 25th, 50th, 75th and 90th percentiles of the distribution as well as the minimum, mean, maximum and standard deviation (SD).

The Cronbach’s alpha reliability index for the eight-item oral reading comprehension sub-task was 0.74, which is within the normally acceptable range of 0.7–0.9 (Tavakol & Dennick 2011), indicating an acceptable level of inter-relatedness. Because the other four of the five subtasks were timed, an alpha index cannot be computed for them. There were moderate to very strong correlations between the various sub-components of reading across the three African languages, as shown in Table 3, suggesting internal consistency reliability

TABLE 2: Descriptive statistics for early-grade reading assessment sub-components by language.

Language	Rcpm	Sample	10th Perc	25th Perc	50th Perc	75th Perc	90th Perc	Min	Mean	Max	SD
Northern Sotho	Total letters rcpm	135	11	19	31	41	48	1	30.4	62	14.0
	Total single words rcpm	135	1	8	22	29	35	0	19.3	40	12.1
	Total non-words rcpm	135	0	4	13	20	25	0	12.5	30	8.9
	Total ORF rcpm	135	0	11	41	55	67	0	36.2	109	25.7
	Story title	132	0	1	2	2	2	0	1.4	2	0.8
	Oral comprehension	135	0	0	1	3	4	0	1.7	6	1.7
Xitsonga	Total letters rcpm	111	5	17	38	48	60	0	34.3	69	19.5
	Total single words rcpm	111	1	9	20	27	31	0	18.2	48	11.5
	Total non-words rcpm	111	1	6	16	22	26	0	14.8	42	9.9
	Total ORF rcpm	111	0	13	47	57	71	0	39.8	91	25.9
	Story title	111	0	2	2	2	2	0	1.6	2	0.8
	Oral comprehension	111	0	2	4	5	6	0	3.4	8	2.2
isiZulu	Total letters rcpm	494	4	13	25	39	49	0	26.4	85	17.1
	Total single words rcpm	494	0	8	19	27	33	0	17.8	44	11.5
	Total non-words rcpm	494	0	5	14	21	25	0	13.5	45	9.2
	Total ORF rcpm	494	0	6	21	33	42	0	21.0	67	15.6
	Story title	488	0	1	2	2	2	0	1.4	2	0.8
	Oral comprehension	494	0	0	2,5	4	5	0	2.4	6	2.0
Total	Total letters rcpm	740	5	15	28	41	51	0	28.3	85	17.2
	Total single words rcpm	740	0	8	20	27	33	0	18.2	48	11.6
	Total non-words rcpm	740	0	5	14	21	26	0	13.5	45	9.3
	Total ORF rcpm	740	0	8	26	42	54	0	26.6	109	21.1
	Story title	731	0	1	2	2	2	0	1.4	2	0.8
	Oral comprehension	740	0	0	2	4	5	0	2.4	8	2.1

Rcpm, read correctly per minute; ORF, oral reading fluency; perc, percentile; SD, standard deviation.

for these adapted versions of the EGRA. Performance in one foundational reading sub-task was related to performance in the other reading subtasks. For example, learners who could only read a few words were also only likely to know a few letter-sounds; conversely, learners who could read more words also knew more letter-sounds.

Some of the notable findings are listed below:

- *Letter-sounds correct per minute*: On the whole, letter-sound knowledge was low, with a mean of 28 letters rcpm. Of the 740 learners assessed, only a quarter of learners could name at least 40 letter-sounds correctly per minute. Across all languages, 25% could only sound out at most 15 letters correctly in 1 min.
- *Word reading*: Word reading, irrespective of orthography, was fairly similar across the three languages, from 22 wcpm in Northern Sotho, 20 wcpm in Xitsonga to 19 wcpm in isiZulu at the 50th percentile. Interestingly, when single-syllable function words typical of the disjunctive orthographies are excluded from a word reading list, learners in Northern Sotho and Xitsonga read at rates similar to learners in isiZulu. Predictably, reading non-words was slower than reading words.
- Although word reading was slow across all the three languages, the reading of words was slightly faster and more accurate than the reading of non-words.
- *Oral reading fluency*: The ORF scores in isiZulu (21 wcpm at the 50th percentile) were considerably lower than those in Northern Sotho (41 wcpm) and Tsonga (47 wcpm). The longer words in written isiZulu texts result in slower reading rates. The occurrence of several short grammatical morphemes that are written separately in the more disjunctive orthographies of Northern Sotho and Xitsonga results in faster reading rates in ORF passages in those languages. (Given that Xitsonga is less disjunctive than

Northern Sotho, the higher mean ORF rates at the 50th percentile in Xitsonga compared to Northern Sotho may be somewhat surprising. Further analysis shows that the disjunctive or conjunctive continuum affects reading rates accordingly.)

- *Oral reading comprehension*: Reading comprehension was generally low, with the median score being 2 out of 8 across the sample.

Fluency and accuracy

Table 4 shows the mean for letter-sounds attempted and the percentage of letters sounded incorrectly. It would seem that while those learners in ORF decile-1 make more errors than those in the higher ORF deciles, almost the entire sample read 15% – 20% of the attempted letter-sounds incorrectly. This low level of letter-sound knowledge and accuracy might be a reflection of early reading instructional practices, where teachers may not be spending time effectively on systematic phonics instruction, especially of the complex consonant system. This result may also reflect lower levels of accuracy in letter-sound reading than in word reading, where words provide a context for the letter-sounds. Perhaps most importantly, those who read the ORF passage at very slow rates (0 wcpm or 1–10 wcpm) also have exceedingly high rates of inaccuracy, making mistakes on every second letter-sound attempted. This is not an insignificant percentage of the sample, accounting for 27% of all learners (202/740). If learners are as likely to get letter-sounds right as wrong, it will be almost impossible for them to read words or connected text with understanding.

Table 5 provides the same information but for the mean number of words attempted by learners in the ORF task, as well as the percentage of words read incorrectly. This is reported for deciles of wcpm in the ORF passage. For example, it shows that the nine Northern Sotho learners in decile-1 (reading at 0→10 wcpm) actually attempted 16 words on average but read half (52%) of these words incorrectly. Across all three language groups, faster readers are more accurate than slower readers. Comparison across the languages shows that accuracy seems to be more important for fluent reading in isiZulu than in Northern Sotho or Xitsonga. The isiZulu learners reading at 21 wcpm or faster read with 95% accuracy or higher. In contrast, 95% accuracy is only associated with

TABLE 3: Correlations between sub-components of reading.

Correlations r	Northern Sotho	Xitsonga	isiZulu
letter-sound × word reading	0.74	0.76	0.60
letter sound × non-word reading	0.69	0.75	0.58
letter-sound × ORF	0.68	0.75	0.55
word reading × non-word reading	0.91	0.92	0.91
word reading × ORF	0.92	0.92	0.91
ORF × comprehension	0.87	0.78	0.81

ORF, oral reading fluency.

TABLE 4: Mean letter-sounds attempted and percentage incorrect by decile of oral reading fluency words correct per minute.

Wcpm	Northern Sotho			Xitsonga			isiZulu		
	Letters attempted	% Incorrect	Sample	Letters attempted	% Incorrect	Sample	letters attempted	% Incorrect	Sample
0 wcpm	11	46	24	10	47	22	11	55	101
Decile 1 (1→10)	21	41	9	7	60	3	20	38	43
Decile 2 (11→20)	26	26	11	21	25	4	26	29	95
Decile 3 (21→30)	27	28	9	32	14	4	29	25	104
Decile 4 (31→40)	35	18	13	34	12	10	34	19	97
Decile 5 (41→50)	37	19	26	39	16	26	42	15	46
Decile 6 (51→60)	41	19	18	44	13	21	40	18	6
Decile 7 (61→70)	37	16	17	43	23	9	43	18	2
Decile 8 (71→80)	36	11	4	57	13	7	-	-	-
Decile 9 (81→90)	48	7	2	56	16	4	-	-	-
Decile 10 (91→100)	43	7	2	53	43	1	-	-	-

Wcpm, words correct per minute.

TABLE 5: Mean oral reading fluency words attempted and percentage incorrect by decile of oral reading fluency words correct per minute.

Wcpm	Northern Sotho			Xitsonga			isiZulu		
	Words attempted	% Incorrect	Sample	Words attempted	% Incorrect	Sample	Words attempted	% incorrect	Sample
Decile 1 (1→10)	16	52	9	6	39	3	9	36	43
Decile 2 (11→20)	21	26	11	19	24	4	18	10	95
Decile 3 (21→30)	34	24	9	29	7	4	26	4	104
Decile 4 (31→40)	40	8	13	36	2	10	36	3	97
Decile 5 (41→50)	49	7	26	47	2	26	46	2	46
Decile 6 (51→60)	57	4	18	55	0	21	53	1	6
Decile 7 (61→70)	67	3	17	63	1	9	68	3	2
Decile 8 (71→80)	75	2	4	76	0	7	-	-	-
Decile 9 (81→90)	86	3	2	87	1	4	-	-	-
Decile 10 (91→100)	107	0	2	91	0	1	-	-	-

Wcpm, words correct per minute.

reading at 51 wcpm or faster in Northern Sotho and 31 wcpm or faster in Xitsonga. One of the reasons why decile-1 learners are reading so slowly is that they are making mistakes on every second or third word. The fastest Northern Sotho readers (wcpm = 107) and Xitsonga readers (wcpm = 91) in the sample made no mistakes whatsoever.

Relationships between letter-sounds, word reading and oral reading fluency

The results show robust and significant correlations between all the sub-components of reading (Table 3). Knowledge of letter-sounds is strongly associated with the ability to read words and non-words, as well as with ORF, although to a slightly lesser degree in the conjunctive reading of isiZulu – an aspect that requires further investigation. Oral reading fluency and comprehension also show a very strong relationship. A more nuanced view of skills in these subtasks and their relationships can clearly be seen in the box plots in Figure 2, showing increasing skill across the deciles.

Figure 2 shows the strong and predictable relationship between both *letters read correctly per minute* and ORF (panels A, C and E), as well as between *single words read correctly per minute* and ORF (panels B, D and F). Decile-0 in the graph represents learners who scored zero on the ORF task; decile-1 represents those scoring 0→10 wcpm; decile-2 those who scored 11–20 wcpm and so on. Looking across the three language groups we can see that approximately 75% of the learners in decile-0 could only pronounce 15 or fewer letter sounds in 1 min and less than five single words in 1 min.

The similarities between Northern Sotho and Xitsonga are clear, particularly when looking at the right panel graphs (single words correct per minute and ORF). There is a tight interquartile range of approximately 5–10 single words per ORF decile. This shows the lock-step relationship between reading single words correctly and connected text fluently.

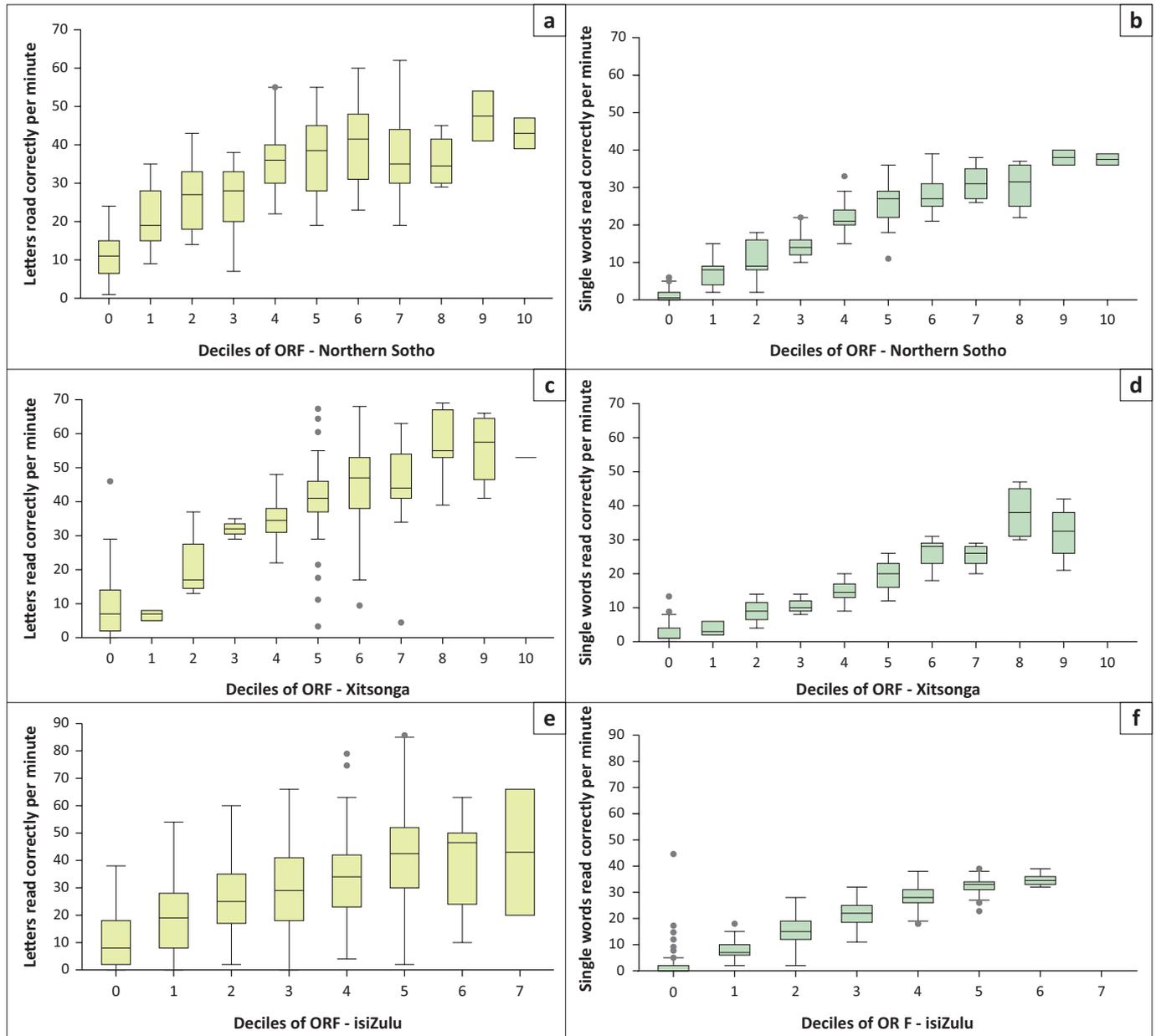
The 'slope' of the right panel graphs is clearly steeper for isiZulu, showing a strong relationship where the interquartile range of single words roughly maps to the ORF decile, that is, for the ORF decile-3 (ORF scores of 20–30 wcpm) the single-word interquartile range is about 19–25. This is in

contrast to both Northern Sotho and Xitsonga which exhibit flatter slopes, that is, these learners are reading fewer single words correct in a minute than ORF words. For example, in Northern Sotho, learners in ORF decile-5 (ORF scores of 40–50 wcpm) are only reading 22–30 single wcpm. While this may initially seem surprising, a closer inspection of the EGRA assessment provides a logical explanation: the single-word assessment included only lexical words and excluded all function words, as explained earlier.

Developing a framework for early reading development in African languages

Setting up reading benchmarks helps establish expectations for reading performance and helps schools identify children in need of support. When developing benchmarks for languages or grades one can take the approach of norming to the population as a whole. For example, Hasbrouck and Tindal (2006:637) collected ORF data from students across the performance spectrum including gifted as well as dyslexic readers. However, this approach becomes problematic in South Africa where the level of reading achievement in the country is so low that any population norms would be unacceptably low. To illustrate, while 96% of American Grade 4 learners reached the Low International Benchmark on PIRLS, only 22% of South African Grade 4 learners reached this rudimentary benchmark (Mullis et al. 2007:69).

If one cannot benchmark to national norms, what are the alternatives? As in earlier work (Draper & Spaul 2015), we argue that benchmarking to comprehension outcomes is a feasible and justifiable alternative. Given that comprehension is the goal of reading, linking reading benchmarks to this outcome seems logical, and this is the approach we take in the present study. Ideally, to this end it is desirable to obtain ORF measures from at least two text passages, and to obtain a comprehension score independently of the comprehension score obtained from the ORF text. However, given the time and budget constraints of the project, the EGRA ORF and oral reading comprehension subtasks were used as a preliminary step in exploring whether there was a comprehension 'tolerance level' in foundational reading skills, that is, whether decoding skills below a certain configuration seriously jeopardised comprehension.



Note: For oral reading fluency eciles 0 = 0 words correct per minute; 1 = 0–10 words correct per minute; 2 = 11–20 words correct per minute; 3 = 21–30 words correct per minute, et cetera. ORF, oral reading fluency.

FIGURE 2: Boxplots of total letters read correctly per minute and total single words read correctly per minute by oral reading fluency deciles.

As part of the adapted EGRA there were eight oral comprehension questions presented to learners after their minute of ORF reading. Using the total scores on these comprehension questions as a rough classification tool, we group learners into one of the four categories: (1) non-readers (those who could not read the title of the story properly), (2) pre-readers (1–2 on comprehension; < 25%), (3) emergent readers (3–4 on comprehension; 26–50%) and (4) basic readers (5+ on comprehension; 62.5%+).

While these are somewhat arbitrary categories, and a short oral comprehension assessment is not ideal as the metric of comprehension, we argue that this is less of a problem for our purposes. Ultimately, we are trying to establish nascent benchmarks for reading letter-sounds, single words,

non-words and connected text for previously unexamined languages. Part of this is identifying the levels of each subcomponent that are typically found together for the same learner. We believe that there is a similar underlying cognitive-linguistic data generating process that is consistent within a language. Our descriptive statistics seem to support this given the relatively narrow range of letter-sound and single-word scores associated with certain ORF deciles. Table 6 shows a similarly narrow interquartile range for ORF scores relative to comprehension categories.

What Table 6 seems to show is that there are certain 'minimum thresholds' below which one cannot find learners that have the requisite comprehension outcomes. To identify these, we look at the 25th percentile score for the emergent

TABLE 6: Early-grade reading assessment subtest distributions by comprehension categories showing median scores with interquartile ranges presented in brackets.

Language	Comprehension Categories	Read correctly per minute				Sample	
		Letters	Single words	Non-words	Connected text (ORF)	<i>n</i>	%
Northern Sotho	Non-readers	24 (17–31)	8 (3–11)	4 (2–8)	14 (7–25)	15	11
	Pre-readers	25 (28–41)	23 (18–28)	14 (9–19)	43 (34–48)	48	36
	Emergent	42 (29–49)	30 (27–35)	21 (18–26)	58 (52–62)	27	20
	Basic	43 (39–46)	33 (27–36)	24 (21–26)	70 (66–84)	12	9
Xitsonga	Non-readers	16 (10–25)	6 (3–11)	6 (4–9)	12 (7–18)	4	4
	Pre-readers	33 (18–41)	16 (13–20)	15 (9–18)	40 (32–50)	16	14
	Emergent	39 (34–48)	19 (15–23)	16 (11–20)	48 (39–51)	30	27
	Basic	46 (38–55)	28 (21–31)	21 (15–26)	57 (48–71)	43	39
isiZulu	Non-readers	19 (9–23)	6 (3–13)	5 (3–9)	4 (1–15)	37	7
	Pre-readers	26 (15–38)	15 (10–20)	11 (7–16)	13 (9–22)	83	17
	Emergent	34 (20–43)	23 (18–29)	17 (13–23)	28 (20–35)	145	29
	Basic	34 (24–48)	30 (26–33)	22 (19–26)	37 (32–43)	102	21

ORF, oral reading fluency; *n*, sample size.

readers category. For example, to get 25% or more on the comprehension questions (emergent-reader) one would need to read at least 53 wcpm in Northern Sotho, 39 wcpm in Xitsonga and 20 wcpm in isiZulu. We will refer to these as the ‘*minimum* fluency thresholds’ for reading in these languages. Interestingly, these figures are very similar to the lowest levels at which learners had 95% accuracy in reading connected text (ORF). These were 51+ wcpm (Northern Sotho), 31+ wcpm (Xitsonga) and 21+ wcpm (isiZulu) – see Table 4. If one takes a more reasonable comprehension metric that learners should achieve 62.5% or more, then learners need to be reading at least 66 wcpm in Northern Sotho, 48 wcpm in Xitsonga and 32 wcpm in isiZulu. We will refer to these as the ‘*minimum* comprehension thresholds’ for reading in these languages. It is important to note that the sample sizes across the languages are not the same, and the numbers of learners in each category are also not equal. While we still think that it is helpful to report the ranges for non-readers and pre-readers for Xitsonga and Northern Sotho, given their small sample sizes we would caution against over-interpreting the results in these categories.

Concluding remarks

The concern about low literacy levels in developing countries such as South Africa is a valid and urgent one. Factors such as reduced time on task, inadequate access to reading materials in African languages and poor quality early reading instruction in high-poverty contexts all contribute to low literacy levels (e.g. De Stefano 2012). In this article, we have probed beneath the comprehension iceberg to better understand how different components of reading promote or hinder reading in agglutinating African languages with transparent disjunctive and conjunctive orthographies. The results show that across all three languages, alphabetic knowledge, accuracy and speed matter in reading. This finding is supported by research into reading in other alphabetic languages elsewhere in the world (Jenkins et al. 2003; Siedenberg 2017). Accuracy and speed were reflected in all the sub-components of the reading test, with a knock-on effect from the most basic reading level, namely, letter-sounds, through word reading to ORF passage reading. Not surprisingly, the learners who demonstrated the highest level

of comprehension were learners who read faster and more accurately than their peers.

The knowledge of letter-sounds showed strong relationships with both word and non-word reading, suggesting that in the early stages, readers in transparent orthographies rely on letter-sound conversion to decode words accurately. Learners who could not sound out, minimally, 25–30 letters correctly per minute on this subcomponent of the test fell into the non-reader or pre-reader categories, suggesting that although they were entering their third year of schooling, they had not yet been launched on a successful reading trajectory. Letter-sound knowledge of the complex consonant system in African languages may help to fine-tune phonological awareness, enabling readers to make finer distinctions at the phonemic level, which in turn improves word processing. The reason why learners in Grade 3 are still struggling to master the alphabetic principle is likely linked to ineffective classroom practices; more research is needed to examine this. Systematic phonics instruction early in the Foundation Phase may help to mitigate this backlog in grasping the alphabetic principle.

Performance was better on the word than on non-word reading tasks, with weaker readers (in the lower percentiles) performing more poorly in the non-word task. Performance on the two-word subtasks was highly correlated. All these three findings resonate with reading research in other agglutinating languages (Miller et al. 2019). Although reading scores did not differ much across languages in the word and non-word subtasks when function words were excluded, large differences in ORF scores showed up when learners read extended text. Differences in word length in the disjunctive and conjunctive orthographies of Northern Sotho and isiZulu, respectively, affect the reading rate. This has important implications for benchmarking and identifying at-risk readers at different grade levels.

Although more research is still needed, the differential reading rates in the conjunctive or disjunctive orthographies have implications for streamlining the benchmarking process; rather than establish benchmarks for each individual African

language (a costly and time-consuming process), sets of benchmarks for the conjunctive and disjunctive orthographies, respectively, may suffice, derived from large-scale systematic and rigorous collection and analysis of performance on EGRA tasks, such as letter sounds, ORF and reading comprehension across grades and language groups. Separate, intermediate benchmarks for languages that show features of both orthographies, such as Xitsonga, should also be established.

Comprehension was compromised when speed and accuracy dropped below the minimum thresholds. Reading below 50 wcpm and 40 wcpm in Grade 3 seems to signal at-risk readers in Northern Sotho and Xitsonga, respectively, while reading below 20 wcpm signals an at-risk reader in isiZulu. If a comprehension threshold of at least 60% is desired, then learners should be reading at least 10 wcpm faster than the above scores in the respective languages. Obviously, more research is needed before benchmarks in different African languages can be established with empirical confidence, but minimum threshold such as these can help teachers develop a sense of what reading success or failure might look like in different languages in the Foundation Phase.

Irrespective of whether languages are analytic or agglutinating, have transparent or opaque scripts, systematic phonics instruction tailored to language-specific orthographic characteristics can provide children learning to read an alphabetic script with letter-sound knowledge that forms accurate building blocks pertinent for word reading in their language. Easy access to reading material will also be critical. Fluency in word and passage reading is built up through daily opportunities to practise reading extended texts in and out of the classroom (National Reading Panel 2000; Spear-Swerling 2006).

It is also important to identify learners who get off to a slow start in reading in the first 3 years of schooling. There is no 'one-size-fits-all' approach; reading benchmarks are specific to languages or language families. In order to reduce inequalities in literacy, it is important for teachers in developing countries to be aware of appropriate reading benchmarks in different languages in which reading is taught. We argue here that we need to move beyond a repetitive focus on low comprehension outcomes; this is the tip of the iceberg. Below the surface, there is widespread evidence that most children have not acquired the basic 'tools' for reading success – the ability to accurately and fluently decode letters and words and move from an effortful activity to an automated skill.

Acknowledgements

This article first appeared as a Research on Socio-Economic Policy (ReSEP) working paper with the same title (series no. WP01/2018) from the Department of Economics, University of Stellenbosch. The data in this article come from the Leadership for Literacy Project funded by the Economic and Social Research Council (ESRC) and the Department for International Development (DFID) of the United Kingdom.

Competing interests

There were no competing interests.

Authors' contributions

E.P. attended to the literature review of the article, N.M. wrote the methodology section and N.S. analysed the data and presented the Results section. All the authors contributed to the Introduction, Discussion and concluding sections.

Funding information

This study is funded by the Economic and Social Research Council (ESRC) and the Department for International Development (DFID) of the United Kingdom under the Raising Learning Outcomes in Education Systems 2015 research grants.

Data availability statement

All data used in the analysis of this paper are now publicly available for researchers to download. This can be done via DataFirst's Open Data Portal. Title "South Africa – Leadership for Literacy 2017" (REF: zaf-resep-lfl-2017-v1). Available: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/726/related_citations.

Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

References

- Acha, J., Laka, I. & Perea, M., 2010, 'Reading development in agglutinative languages: Evidence from beginning, intermediate and adult Basque readers', *Journal of Experimental Child Psychology* 105(4), 359–375. <https://doi.org/10.1016/j.jecp.2009.10.008>
- Adams, M., 1990, *Beginning to read: Thinking and learning about print*, MIT Press, Cambridge, MA.
- Adams, M.J., 1994, 'Modeling the connections between word recognition and reading', in R.B. Ruddell, M.R. Ruddell & H. Singer (eds.), *Theoretical models and processes of reading*, pp. 838–863, International Reading Association, Newark, NJ.
- Aro, M. & Wimmer, H., 2003, 'Learning to read: English in comparison to six more regular orthographies', *Applied psycholinguistics* 24(4), 621–635. <https://doi.org/10.1017/S0142716403000316>
- Babayağit, S. & Stainthorp, R., 2007, 'Preiterate phonological awareness and early literacy skills in Turkish', *Journal of Research in Reading* 30(4), 394–413. <https://doi.org/10.1111/j.1467-9817.2007.00350.x>
- Blachowicz, C.L.Z., Moskal, M.K., Massarelli, J.R., Obrochta, C.M., Fogelberg, E. & Fisher, P., 2006, "'Everybody reads': Fluency as a focus for staff development", in T. Rasinski, C. Blachowicz & K. Lems (eds.), *Fluency instruction: Research-based practices*, pp. 141–154, The Guilford Press, New York, NY.
- Blaiklock, K., 2004, 'The importance of letter knowledge in the relationship between phonological awareness and reading', *Journal of Research in Reading* 27(1), 36–57. <https://doi.org/10.1111/j.1467-9817.2004.00213.x>
- Burgess, S.R. & Lonigan, C.J., 1998, 'Bi-directional relations of phonological sensitivity and pre-reading abilities: Evidence from a pre-school sample', *Journal of Experimental Child Psychology* 70(2), 117–141. <https://doi.org/10.1006/jecp.1998.2450>
- Castles, A., Rastle, K. & Nation, K., 2018, 'Ending the reading wars: Reading acquisition from novice to expert', *Psychological Science in the Public Interest* 19(1), 5–51. <https://doi.org/10.1177/1529100618772271>
- Cummings, K.D. & Petcher, Y. (eds.), 2016, *The fluency construct: Curriculum-based measurement, concepts and applications*, Springer, New York.
- De Stefano, J., 2012, *State of the art knowledge in education: Opportunities to learn*, USAID/EQUIP2, FHI360, Washington, DC.
- De Vos, M., Van der Merwe, K. & Van der Mescht, C., 2014, 'A linguistic research programme for reading in African languages to underpin CAPS', *Journal for Language Teaching* 48(2), 149–177. <https://doi.org/10.4314/jlt.v48i2.7>

- Diemer, M.N., 2015, 'The contributions of phonological awareness and naming speed to the reading fluency, accuracy, comprehension and spelling of Grade 3 isiXhosa readers', unpublished MA dissertation, Rhodes University, Grahamstown.
- Draper, K. & Spaul, N., 2015, 'Examining oral reading fluency among Grade 5 rural ESL learners in South Africa: An analysis of NEEDU 2013', *South African Journal of Childhood Education* 5(2), 44–77. <https://doi.org/10.4102/sajce.v5i2.382>
- Ehri, L.C., 2005, 'Learning to read words: Theory, findings and issues', *Scientific Studies of Reading* 9(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4
- Ellis, N.C. & Hooper, M., 2001, 'Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency matched tests', *Applied Psycholinguistics* 22(4), 571–599. <https://doi.org/10.1017/S014216401004052>
- Fuchs, L.S., Fuchs, D., Hosp, M.K. & Jenkins, J.R., 2001, 'Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis', *Scientific Studies of Reading* 5(3), 239–256. https://doi.org/10.1207/S1532799XSSR0503_3
- Guthrie, J.T., Hoa, A.L.W., Wigfield, A., Tonks, S.M., Humenick, N.M. & Littles, E., 2007, 'Reading motivation and reading comprehension growth in the later elementary years', *Contemporary Educational Psychology* 32(3), 282–313. <https://doi.org/10.1016/j.cedpsych.2006.05.004>
- Hasbrouck, J. & Tindal, G.A., 2006, 'Oral reading fluency norms: A valuable assessment tool for reading teachers', *The Reading Teacher* 59(7), 636–644. <https://doi.org/10.1598/RT.59.7.3>
- Howie, S., Combrinck, C., Roux, K., Tshela, M., Mokoena, G.M. & McLeod Palane, N., 2017, 'PIRLS LITRACY 2016: South African Highlight Report', Centre for Evaluation and Assessment, University of Pretoria, Pretoria.
- Howie, S., Van Staden, S., Tshela, M., Dowse, C. & Zimmerman, L., 2012, 'PIRLS 2011: Progress in International Reading Literacy Study 2011: South African children's reading literacy achievement: summary report', Summary Report, Centre for Evaluation and Assessment, University of Pretoria, Pretoria.
- Howie, S., Venter, E., Van Staden, S., Zimmerman, L., Long, C., Du Toit, C. et al., 2006, 'Progress in International Reading Literacy Study 2006', Centre for Evaluation and Assessment, University of Pretoria.
- Jenkins, J.R., Fuchs, L.S., Van den Broek, P., Espin, C. & Deno, S.L., 2003, 'Sources of individual differences in reading comprehension and fluency', *Journal of Educational Psychology* 95(4), 719–729. <https://doi.org/10.1037/0022-0663.95.4.719>
- Jimerson, S.R., Hong, S., Stage, S. & Gerber, M., 2013, 'Examining oral reading fluency trajectories among English language learners and English speaking students', *New Approaches in Educational Research* 2(1), 3–11. <https://doi.org/10.7821/naer.2.1.3-11>
- Kim, Y.-S.G. & Piper, B., 2019, 'Component skills of reading and their structural relations: Evidence from three sub-Saharan African languages with transparent orthographies', *Journal of Research in Reading* 42(2), 1–23. <https://doi.org/10.1111/1467-9817.12271>
- Lekgoko, O. & Winskel, H., 2008, 'Learning to read in Setswana and English: Cross-language transference of letter-knowledge, phonological awareness and word reading skills', *Perspectives in Education* 26(4), 57–73.
- Makaure, Z.P., 2017, 'Phonological processing and reading development in Northern-Sotho – English bilinguals', Unpublished MA dissertation, University of South Africa, Pretoria.
- Miller, P., Guldenoglu, B. & Kargin, T., 2019, 'Reading failure in a completely transparent orthography representing a morphologically highly complex agglutinative language: The case of Turkish', *Journal of Developmental and Physical Disabilities* 31(5), 669–689. <https://doi.org/10.1007/s10882-019-09667>
- Miller, P., Kargin, T. & Guldenoglu, B., 2014, 'Differences in the reading of shallow and deep orthography: Developmental evidence from Hebrew and Turkish readers', *Journal of Research in Reading* 37(4), 409–432. <https://doi.org/10.1111/j.1467-9817.2012.01540.x>
- Mullis, I., Martin, M., Kennedy, A. & Foy, F., 2007, *PIRLS 2006 International Report: The Progress in International Reading Literacy Study in Primary Schools in 40 Countries*, TIMSS & PIRLS International Study Center, Boston, MA.
- Muter, V. & Diethelm, K., 2001, 'The contribution of phonological skills and letter knowledge to early reading development in a multilingual population', *Language Learning* 51(2), 187–219. <https://doi.org/10.1111/1467-9922.00153>
- National Reading Panel, 2000, *Teaching children to read: An evidence-based assessment of the scientific literature on reading and its implications for reading instruction*, US Government Printing Office, Washington, DC.
- Nieto, J.E., 2005, 'Analysis of difficulties in understanding and applying the alphabetic principle', *Journal of Research in Educational Psychology* 2(2), 75–104.
- Oney, B. & Durgunoglu, A.Y., 1997, 'Beginning to read in Turkish: A phonologically transparent orthography', *Applied Psycholinguistics* 18(1), 1–15. <https://doi.org/10.1017/S01421640000984X>
- Perfetti, C.A., 2007, 'Reading ability: Lexical quality to comprehension', *Scientific Studies of Reading* 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>
- Piper, B. & Zuilkowski, S.S., 2016, 'Role of timing in assessing oral reading fluency and comprehension in Kenya', *Language Testing* 31(1), 75–98. <https://doi.org/10.1177/0265532215579529>
- Pretorius, E.J., 2015, 'Failure to launch: Matching language policy with literacy accomplishment in South African schools', *International Journal for the Sociology of Language* 234, 47–76. <https://doi.org/10.1515/ijsl-2015-0004>
- Pretorius, E.J., 2018, *Reading in the African languages: An annotated bibliography 2004-2017*, viewed 06 October 2019, from <https://www.jet.org.za/clearinghouse/printed/resources/language-and-literacy-resources>.
- Pretorius, E.J. & Knoetze, J.J., 2012, 'The teachers' book club: Broadening teachers' knowledge and building self-confidence', *Musaion* 31(1), 27–46.
- Pretorius, E.J. & Spaul, N., 2016, 'Exploring relationships between oral reading fluency and reading comprehension amongst English second language readers in South Africa', *International Journal: Reading and Writing* 29(9), 1449–1471. <https://doi.org/10.1007/s11145-016-9645-9>
- Ouellette, G.P. & Haley, A., 2013, 'One complicated extended family: The influence of alphabetic knowledge and vocabulary on phonemic awareness', *Journal of Research in Reading*, 36(1), 29–41. <https://doi.org/10.1111/j.1467-9817.2010.01486.x>
- Rees, S.A., 2016, 'Morphological awareness in readers of isiXhosa', unpublished MA dissertation, Rhodes University, Grahamstown.
- Richards, J.C. & Schmidt, R., 2010, *Longman dictionary of language teaching and applied linguistics*, Longman, London.
- Share, D.L., 1995, 'Phonological recoding and self-teaching: Sine qua non of reading acquisition', *Cognition*, 55(2), 151–218. [https://doi.org/10.1016/0010-0277\(94\)00645-2](https://doi.org/10.1016/0010-0277(94)00645-2)
- Siedenberg, M., 2017, *Language at the speed of sight: How we read, why so many can't, and what can be done about it*, Basic Books, New York.
- Silven, M., Poskiparta, E. & Niemi, P., 2004, 'The odds of becoming a precocious reader of Finnish', *Journal of Educational Psychology* 96(1), 152–164. <https://doi.org/10.1037/0022-0663.96.1.152>
- Snow, C.A. & Kim, Y.-S., 2007, 'Large problem spaces: The challenge of vocabulary for English language learners', in R.K. Wagner, A.E. Muse & K.R. Tannenbaum (eds.), *Vocabulary acquisition: Implications for reading comprehension*, Guilford Press, New York.
- Spaul, N. & Pretorius, E., 2019, 'Still falling at the first hurdle: Examining early grade reading in South Africa', in N. Spaul & J. Jansen, J (eds.), *South African schooling: The enigma of inequality*, Springer, Cham.
- Spear-Swerling, L., 2006, 'Children's reading comprehension and oral reading fluency in easy text', *Reading and Writing: An Interdisciplinary Journal* 19, 199–220. <https://doi.org/10.1007/s11145-005-4114-x>
- Stanovich, K.E., 1986, 'Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy', *Reading Research Quarterly* 21, 360–406.
- Tavakol, M. & Dennick, R., 2011, 'Making sense of Cronbach's alpha', *International Journal of Medical Education* 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8df>
- Taylor, N. & Taylor, S., 2013, 'Teacher knowledge and professional habitus', in N. Taylor, S. Van den Berg & T. Mabogoane (eds.), *Creating effective schools: Report of South Africa's national schools effectiveness study*, Pearson, Cape Town, pp. 201–232.
- Taylor, S., Cilliers, J., Prinsloo C., Fleisch B. & Reddy V., 2017, 'The early grade reading study: Impact evaluation after two years of interventions', Technical report, Department of Basic Education, Pretoria.
- Thal, D.J., Bates, E., Goodman, J. & Jahn-Samilo, J., 1997, 'Continuity of language abilities. An exploratory study of late- and early-talking toddlers', *Developmental Neuropsychology* 13(3), 239–273. <https://doi.org/10.1080/87565649709540681>
- Wills, G., 2017, *What do you mean by "good"?* The search for exceptional primary schools in South Africa's no-fee school system', Stellenbosch Economic Working Papers, WP 16/2017, Stellenbosch.
- Wilsenach, C., 2013, 'Phonological skills as predictor of reading success: An investigation of emergent bilingual Northern Sotho/English learners', *Per Linguam: a Journal of Language Learning/Tydskrif vir Taalaanleer* 29(2), 17–32. <https://doi.org/10.5785/29-2-554>
- Wilsenach, C., 2015, 'Receptive vocabulary and early literacy skills in bilingual Northern Sotho-English bilinguals', *Reading and Writing* 6(1), Art. #77, 11 pages. <https://doi.org/10.4102/rw.v6i1.77>
- Wise, J.C., Sevcik, R.A., Morris, R.D., Lovett, M.W. & Wolf, M., 2007, 'The growth of phonological awareness by children with reading disabilities: A result of semantic knowledge or knowledge of grapheme-phoneme correspondences?', *Scientific Studies of Reading* 11(2), 151–164. <https://doi.org/10.1080/10888430709336557>
- Wolf, M. & Katzir-Cohen, T., 2001, 'Reading fluency and its intervention', *Scientific Studies of Reading* 5(3), 211–239. https://doi.org/10.1207/S1532799XSSR0503_2
- Ziegler, J. & Goswami, U., 2005, 'Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory', *Psychological Bulletin* 131(1), 3–29. <https://doi.org/10.1037/0033-2909.131.1.3>