

Literacy and numeracy assessment tools for South African children in Grade R and Grade 1

Developing or adapting early mathematics and early literacy assessments for the purpose of the TIME Longitudinal Study

This brief was written for Wordworks by Dr Magali von Blottnitz, with input from Dr Shelley O'Carroll for Wordworks and Cally Kuhne for RED INK. It can be referenced as follows: von Blottnitz, M. (2025). Exploring the implementation of the TIME home-learning programme and learning trajectories of 5-7 year olds, Brief 5, Wordworks: Cape Town.

This is the fifth in a [series of learning briefs](#) that unpack learnings from a longitudinal study into the TIME home-learning programme and the learning trajectories of 5- to 7-year-olds. This brief focuses on the two assessment tools that were used in the study: the WELA (early literacy) and MELA (early mathematics). The brief presents the tools within the greater context of existing tools for that age band, explains briefly how they were developed or adapted, and what lessons were gained from using them. A more comprehensive report is available from Wordworks on request.



This brief seeks to address the following questions:

- Which assessment tools were available at the time of the TIME study and why was it necessary to develop a new tool for mathematics and adapt an existing tool for language and literacy?
- Which principles underpinned the redevelopment of Wordworks' Early Literacy Assessment ("WELA 2.0"), and how was it developed, piloted and reviewed?
- What process enabled the partial development of RED INK's Mathematics Early Learning Assessment ("MELA"), and why was the development left incomplete?
- What did the piloting of the tools and analysis of results teach us about the tools themselves, their usefulness for the sector, and possible pitfalls?
- What could be the way forward for WELA and MELA?

Existing early assessment tools in 2022

The TIME longitudinal study was launched at the beginning of 2022 with the aim (among other objectives) of studying the learning trajectories of children throughout Grade R and Grade 1 in language and mathematics. The study covered Western Cape children schooled in Afrikaans, English and isiXhosa.

Figure 1 summarises the main assessment tools, which have been used in South Africa, either for language and literacy or for mathematics.

Figure 1 – Main assessment tools for early language and literacy, in use in South Africa, for children aged 4-8

Domain	Pre-Grade R	Grade R	Grade 1	Grade 2 +
Includes language and literacy as well as numeracy	ELOM 4&5	ELOM R		
	IDELA			
			TARL Assessment Tasks	
			ANAs (discontinued)	
		JET Customs Assessments		
Language and literacy		MAIN (Narrative Assessment)		
		Click Letter Knowledge Test		
		WELA		
		DIBELS (fluency measures suitable for screening/ benchmarking)		
			EGRA (Reading)	
Mathematics		Marko-D		
			EGMA (Junior EGMA Hybrid)	
			VASSI	

Source: adapted from various sources including Moloi et al (2023) and O'Carroll (2024)¹. See background report for full details.

¹ Moloi et al. (2023), Assessments Learning Brief 2023, Kelello; O'Carroll, S. (2024), *Mapping out early language and literacy assessment tools in South Africa for children aged 3 to 8 years*, 11 April 2024.



Many of the assessment tools listed in Figure 1 are either only available from Grade 1, or cover only the years up to Grade R. This is not surprising, considering that the acquisition of basic reading and writing skills in Grade 1 opens more time- and cost-effective assessment modalities, than what is available in an ECD setting. However, subjecting South African Grade 1 learners to assessments designed for literate children poses several problems:

- Children with delayed acquisition of literacy skills may be unable to perform the tasks
- Children's anxiety and unfamiliarity with test situations may become confounding factors in the interpretation of results
- In low performing schools, floor effects will limit insight into the range of children's abilities
- Splitting pre-school and formal school assessments may impede the understanding of the pivotal years of Grade R-Grade 1.

Since the TIME study's objective was to explore and document the critical early literacy and early mathematical skills that develop in children between the ages of 5 and 7 years, there was a need for consistent tools that span over those two years and target the very competencies applicable to that age. For this effect, the following tools were available:

Table 1 – Assessment tools that span across Grade R and Grade 1

Domain	Assessment	Areas covered and further details	
Literacy	Click test	Letter knowledge only	
	DIBELS	A set of short (one-minute) fluency measures that can be used for universal screening, benchmark assessment, and progress monitoring (Kindergarten to Grade 8).	
	MAIN	Narrative production	
	WELA	<ul style="list-style-type: none"> • Listening comprehension • Drawing and emergent writing • Letter-sound knowledge • Phonological awareness (focus sound, blending) • Reading of words • Writing of words, sentences, story 	
Mathematics	Marko-D	Number sense, with child skills organised along five levels: <ol style="list-style-type: none"> 1. Counting, knowledge of numbers 2. Mental number line 3. Cardinality and decomposability 4. Class inclusion and embeddedness 5. Relationality² 	
Literacy and Mathematics	PAL-ELANA (under development in 12 countries, not including South Africa)	Literacy: <ul style="list-style-type: none"> • Oral language • Symbol (letter) knowledge • Word decoding, encoding and comprehension • Sentence reading and comprehension • Text reading and comprehension 	Mathematics: <ul style="list-style-type: none"> • Data • Patterns • Measurement • Geometry • Numbers

Based on this review, in 2022 the TIME study team decided to use the following assessment tools:

- For literacy, WELA appeared to be the most suitable, even though it was going to require some revisions
- For mathematics, in order to generate insights of comparable quality to WELA, RED INK offered to develop an in-house tool for Grade R and 1, of similar breadth of scope as the WELA. This tool was called MELA.

Hence, a side project was initiated within the TIME study, which involved adapting the WELA tool and developing an equivalent tool for mathematics.



² See Herholdt (2017), p. 15-20

The development and piloting of the revised Wordworks Early Literacy Assessment (WELA 2)

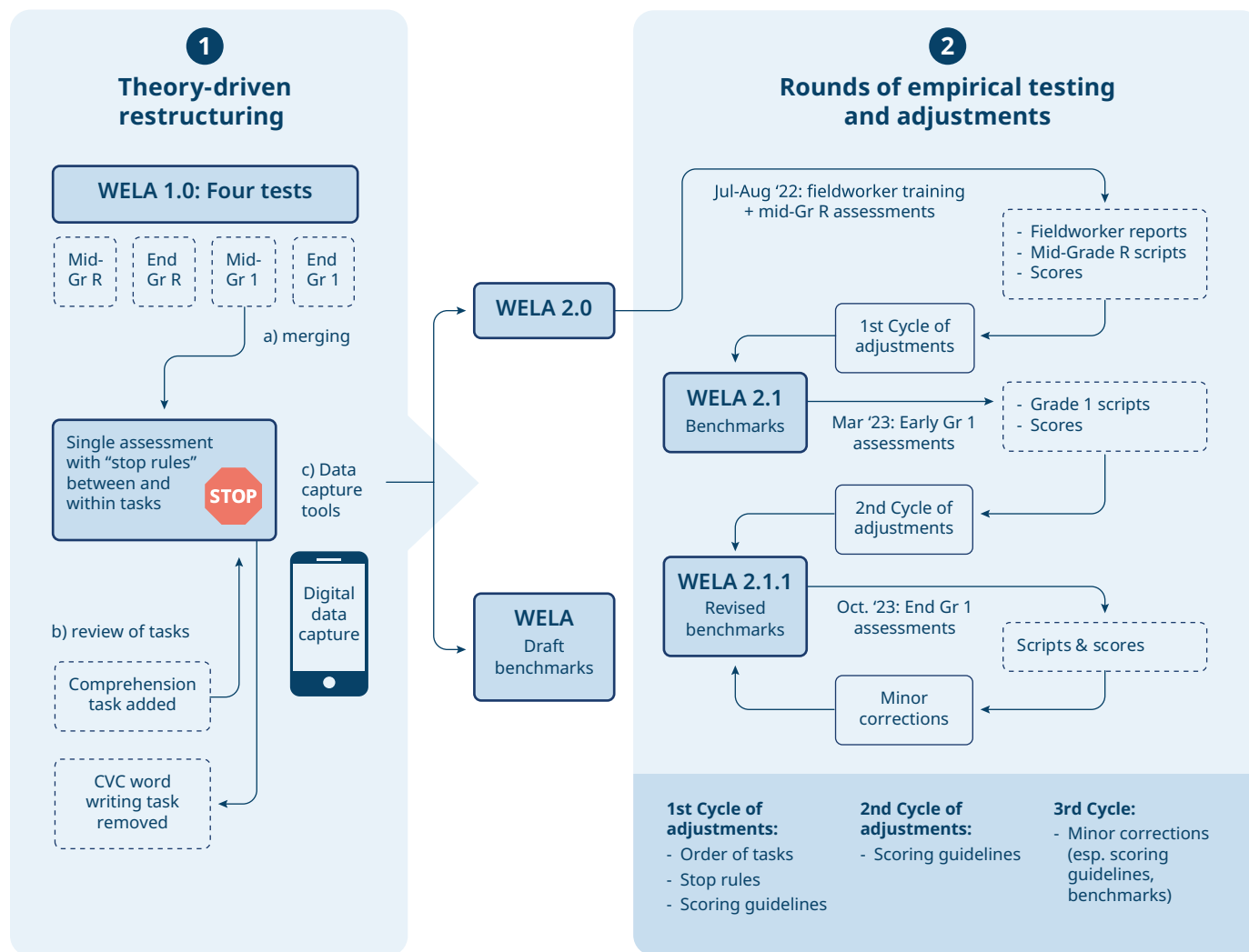
FROM WELA 1 TO WELA 2

Wordworks developed the initial version of its Early Literacy Assessment in 2011 to assist with the targeting of children for their Ready Steady Read Write (RSRW) tutoring programme. The tool was designed for children who were struggling to keep up with the expectations of their grade. It included four discrete and distinct assessments to be administered at roughly 6-monthly intervals between the middle of Grade R and the end of Grade 1. However, in 2022, there was a sense that the tool needed some revisions, including:

- repeating more tasks, in order to enable comparison of task scores over time
- adding a task with a focus on comprehension
- automating the capture and computation of scores through digital technology.

The journey from WELA 1.0 to WELA 2.1.1 involved a number of steps. First, a theory-driven adjustment saw the four discrete tests merged into one single assessment with stop rules. A comprehension task was added but, to prevent excessive length, a word writing task was removed. A digital capture tool was developed on Kobo with automated data validation and computation. The fieldwork sprints of the TIME study were then used to subject this new assessment to multiple iterations of reviews and adjustments, as represented in Figure 2 to the right.

Figure 2 – From WELA 1.0 to WELA 2.1.1: an iterative process

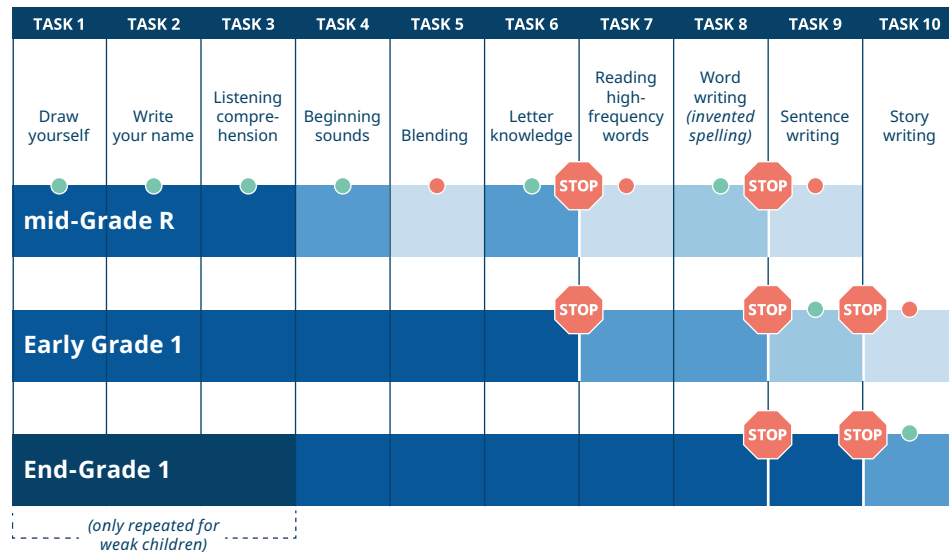


FEATURES OF THE LATEST VERSION OF WELA

WELA tasks

Through its stop rules, WELA becomes a flexible assessment that extends to various degrees of difficulty depending on the child's comfort with the tasks. Figure 3 illustrates, for each assessment intervals, which tasks were done by all, some, or only a small minority of children.

Figure 3 – Appropriateness of WELA tasks according to the child's grade



- Most children able to complete task with acceptable scores
- Most children attempt task, with scores ranging from poor to good
- Only the stronger children qualify for that task
- Only a small minority of children can attempt task
- Task very easy for most children, skipped if repeat assessment
- Task appropriate for that age group
- Task beyond reach of most

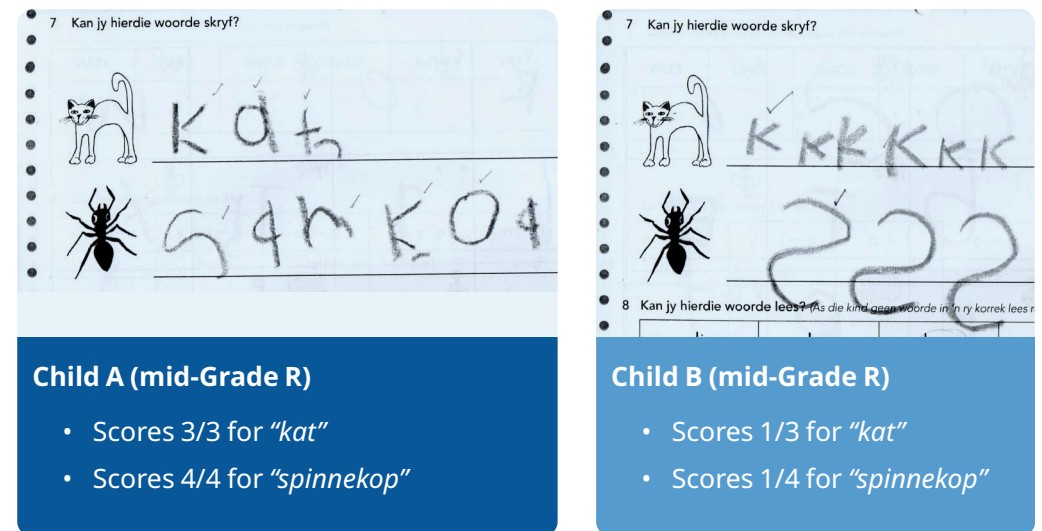
Spirit of scoring

The WELA aims to identify and appreciate the early signs of literacy development, so when scoring, the focus is on seeking evidence of a child's acquisition of some key principles. Where the EGRA puts the emphasis on reading speed and considers a child's answer as either right or wrong, the WELA steers the users towards a better understanding of emergent literacy.

To illustrate this, Figure 4 shares examples of the word-writing task (task 8) in Afrikaans. Learners were requested to have a go writing two words: a simple word, "kat" (cat) and a long and difficult one, "spinnekop" (spider). The task is trying to surface the child's ability to hear at least some of the sounds of each word and transcribe them into the corresponding letters. Since the focus is on the alphabetic principle, a child is not penalised for forming reversed letters or for spelling mistakes.

Child A wrote "kat" correctly; he also identified and transcribed most of the sounds of the word "spinnekop" (although the letter "p" was reversed), which is enough for him to earn full marks in this task. Child B correctly identified and transcribed the initial sound in both words (even though the 's' is reversed), and understood that a word is made of more than one letter, but had not identified or transcribed further sounds in the word. He scores only one mark per word.

Figure 4 – Examples of scoring of the word-writing task (Afrikaans)



FOR DATA INTERPRETATION: THE WELA COMPOSITE SCORE, BENCHMARKS AND TOOLS

Composite score

In its initial version, the WELA didn't include a total score: rather, children's scores in individual tasks would guide the literacy programme on the extent and nature of support that a child would require. However, with the growing interest in using the assessment for M&E purposes, the need for a single summary score became more pressing. However, the variable scope of the assessment, the unequal predictive value of tasks, and the different scoring scales, made it difficult for users to summarise a child's performance into a single indicator.

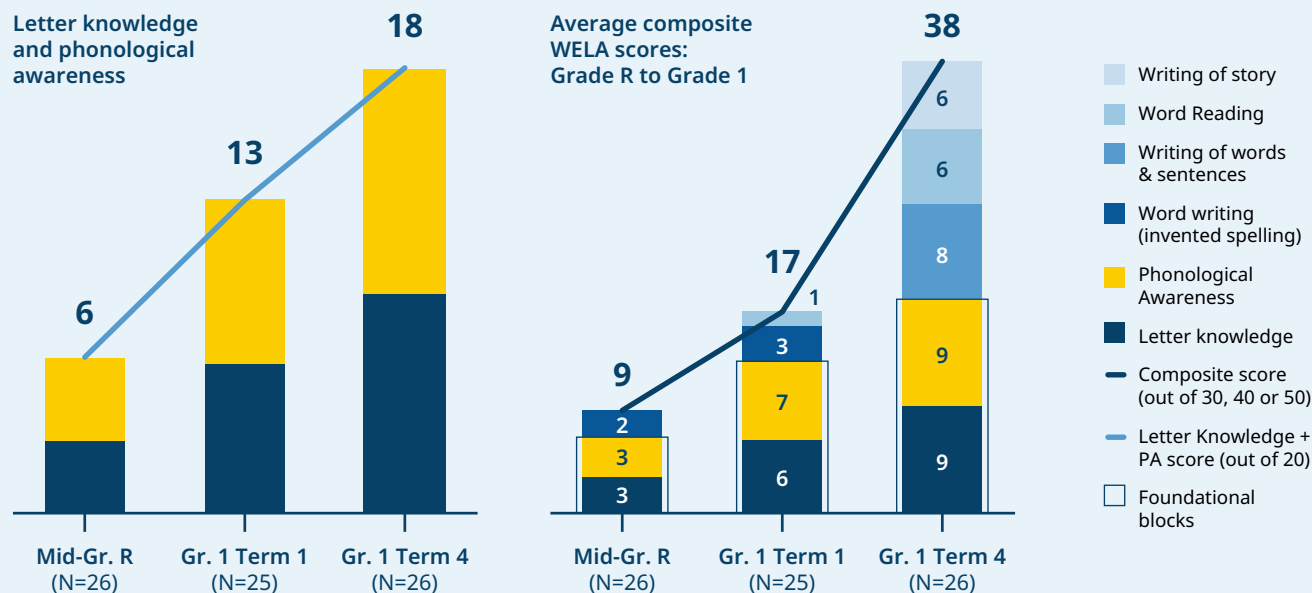
For these reasons, it felt necessary to introduce a composite score, which would allow for comparisons of children's achievement in the main components of the assessment, over time and across a cohort. The technical report associated with this learning brief describes in greater detail the decisions that were taken in building the normed score, which is made up of:

- Two key building blocks of early literacy: letter knowledge and phonological awareness
- Further reading and writing tasks, to the extent that the child was confronted with them (i.e. if the child fulfilled the conditions defined in the 'Stop rules').

Depending on how many tasks a child will complete, their total composite would be out of a total ranging from 30 (early in Grade R) to 50 (end of Grade 1). In other words, a child's composite score will increase through the combined effect of more tasks being accessible to the child and better achievement in repeated tasks.

Figure 5 illustrates how, in the TIME study, the normed score for foundational skills (letter knowledge and phonological awareness) grew from an average of 6/20 at the middle of Grade R to a mean achievement of 18/20 at the end of Grade 1. The mean composite score grew from 9 / 30, to 38 / 50 over the same period³.

Figure 5 – Foundational skills and composite score, Grade R to Grade 1, TIME study



Source: analysis of TIME study assessment scores.

Benchmarks

In 2022-23, a team of early literacy experts proposed a first set of draft benchmarks for WELA. For each task, each language, and at four different points in time, score intervals were proposed for the categories "at risk", "still developing", "on track" or "beyond expectations". In the absence of sufficient representative WELA data for psychometric norming to take place, these benchmarks were based on the developers' extensive knowledge of CAPS, international and South African literature, as well as decades of experience on the field.

Benchmarks can be very helpful if applied to a dataset that has been cleaned thoroughly and is well understood; however, they can also be damaging if used without a sound understanding of the data and its meaning for the children assessed. To mitigate this risk, Wordworks only provides the benchmarks to users on request.

³ It is important to remember that the assessment was designed to avoid ceiling effects; therefore, scores are not normed against a set of minimum expectations for the grade. In other words, the fact that children achieve less than 50% in the tasks they were assessed in does not mean that they are falling behind.

Visualisation tools

Another useful tool for users of WELA is the comprehensive WELA Dashboard developed by JumpTrak™, which provides an overview of progress throughout the year, individual topic and task analysis as well as educator tools for AI assisted action plans for the class based on the WELA assessment results.

Figure 6 – JumpTrak™ Overview for 2024



Source: JumpTrak™ Mellon training guide, 2024

CONCLUSION ON WELA: STRENGTHS, LIMITATIONS AND POTENTIAL

The advantages of WELA are that it is easily accessible at a low cost, has a broad scope and an emerging focus. However, the tool is not suitable for any type of child assessment, due to two groups of limitations. By design, it is time-consuming to administer and complex to score, which affect its usefulness in very large-scale studies. Also, due to resource constraints, it has not (yet) been possible to develop, standardise and norm the instrument in line with internationally acceptable standards for assessment tools.

Despite these limitations, the tool possesses some undeniable strengths and has the potential to play a meaningful role in the education sector in South Africa, as summarised in the table to the right.

Table 2 – Limitations, strengths and potential of the WELA tool

WELA STRENGTHS

- **No restriction** on who can use the tool (only a short training is required)
- Focus on both **writing** and **reading** (beyond mechanics of writing: children's ability to represent their ideas)
- Developmental / **emergent focus** (helps teachers/tutors to track development from mid-Grade R to end Grade 1)
- **Early detection** of children at risk

WELA LIMITATIONS

Design-related limitations:

- Broad scope of tasks => time-consuming to administer
- Attention to nuance => Scoring can be complex

Limited development:

- It has not been tested statistically across a large representative sample:
 - Benchmarks are derived / extrapolated from CAPS
 - No robust psychometric validation yet
- No measure of reading comprehension yet
- Not yet available in all South African languages

WELA POTENTIAL

- Potential to become a common M&E tool for South African literacy NGOs: a meta-database of real-life assessments across a representative sample would help strengthen the tool's validity
- Potential for teachers to use single tasks (e.g. emergent / early writing tasks)
 - to get a snapshot of children's development at different stages,
 - to group children for small group teaching.

The development of the Mathematics Early Learning Assessment (MELA)

THE PROCESS OF DEVELOPING AND REVIEWING MELA FOR GRADE R

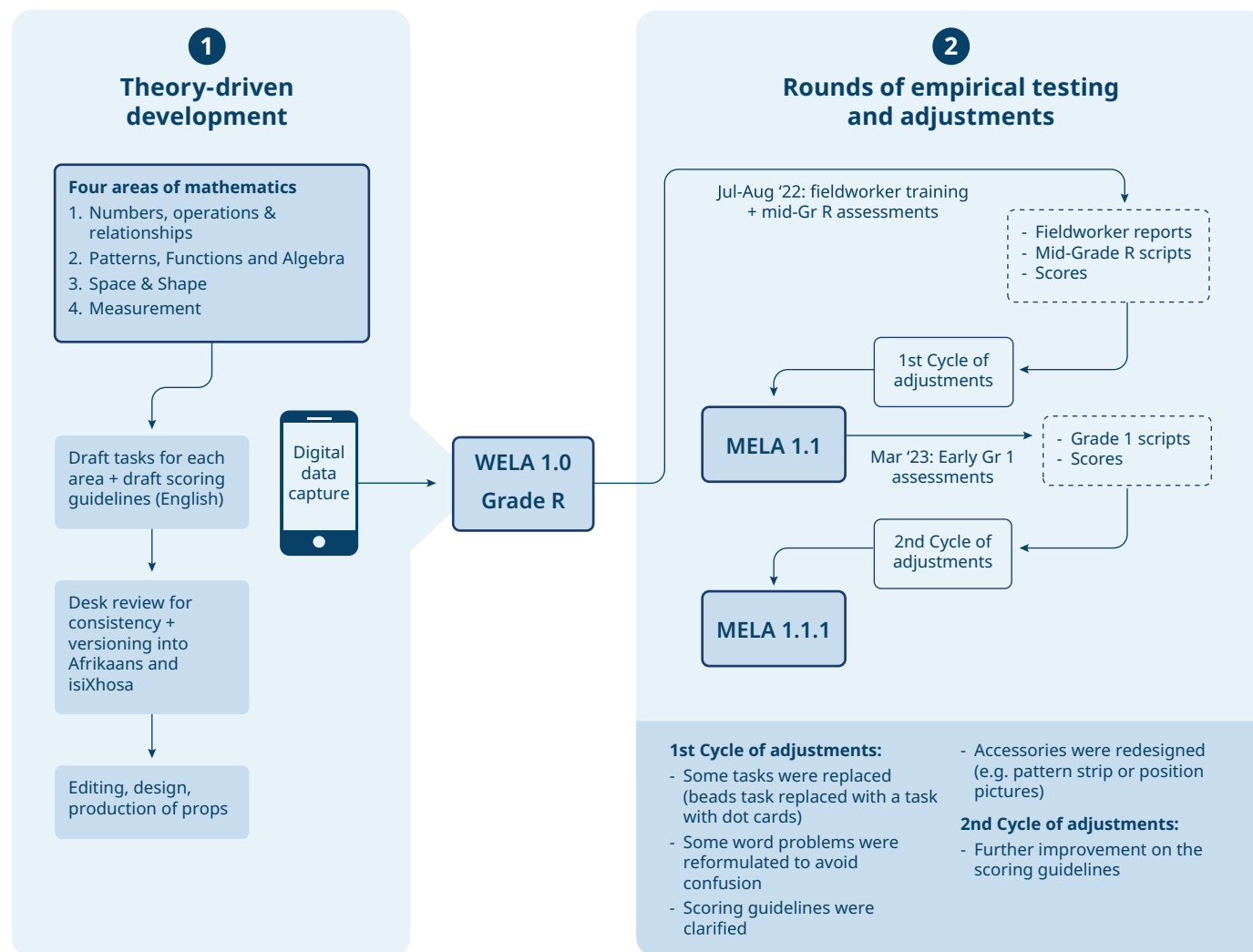
The MELA tool was developed in a joint project set up between Wordworks and RED INK, with tight timeframes due to the demands of the TIME study. RED INK led the conceptual development of the assessment tasks whereas Wordworks took over the versioning, editing, graphic design and digital score capture. Figure 7 represents the process of development and reviews of MELA Grade R.

SCOPE OF MELA: DOMAINS INCLUDED

Unlike many other assessment tools that focus only on counting and operations, but similarly to the WELA, the MELA instrument aims to provide an all-round overview of competencies that develop in early years with regard to mathematical concepts covered in CAPS. Hence, it includes all five content areas of mathematics that apply in the GET band (see DBE 2011), namely:

- **Numbers, Operations and Relationships**, which form the bulk of the assessment
- **Patterns, Functions and Algebra**
- **Space and Shape** (Geometry)
- **Measurement**
- A small component of **Data Handling**, through the sorting of shapes.

Figure 7 – Development, testing and adjustment of the MELA Grade R instrument



MELA FOR GRADE R: KEY FEATURES

Similarly to the WELA, the MELA is designed to be usable by anyone who has some experience of working with young children and does not require any special qualifications. MELA also rests on a developmental approach, and the assessment manual includes several concrete and semi-concrete resources such as dot cards or pictures.

Figure 8 – Selected pages from the MELA manual as well as concrete accessories used for some MELA tasks



MELA FOR GRADE 1: STILL OUTSTANDING

Sadly, hopes to secure some dedicated funding to enable the further development of MELA, and especially the Grade 1 assessment, have not been realised in time for the rounds of fieldwork that took place in 2023 in the context of the TIME longitudinal study. In the absence of such funding, RED INK has not been able to allocate the necessary resources to finalise the Grade 1 assessment.

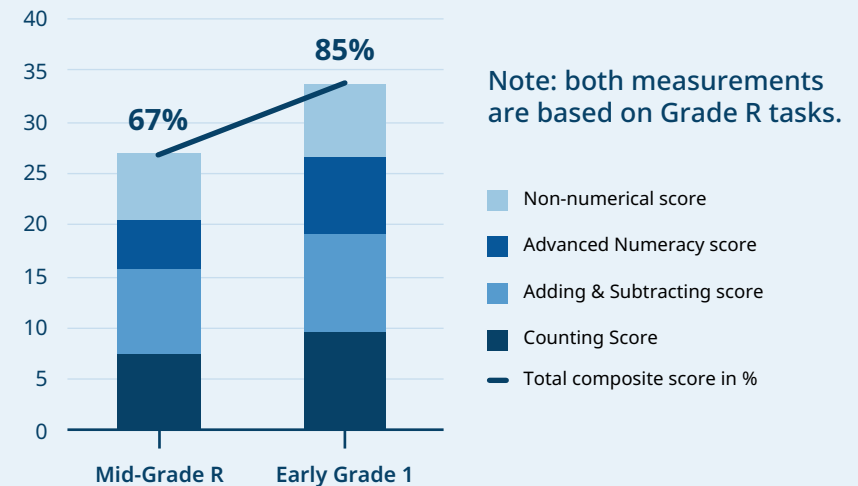
MELA COMPOSITE SCORE

As for the WELA, a composite score was developed to synthesise a child's achievement and make comparisons between children or groups of children easier. It groups selected task scores into the following four domains, each receiving a score out of 10:

1. Counting (oral counting as well as counting of objects)
2. Adding and Subtracting (using either counters, or pictures)
3. Advanced numeracy (which involves word problems as well as sharing and grouping tasks)
4. Non-numerical mathematics (patterns, shapes, positions and measurement)

For illustrative purposes, the graph below represents how the composite score plays out in a sample of 26 Western Cape children between middle of Grade R and beginning of Grade 1.

Figure 9 – Mean MELA composite score achieved in the TIME Study sample, 2022-2023



Source: TIME Longitudinal study

Usefulness and limitations of these tools, and way forward

The piloting process confirmed that the MELA and WELA offer a valuable alternative to other existing tools, if the pitfalls are carefully managed.

VALUE OF BROAD MULTI-COMPETENCY ASSESSMENTS

Although broad assessments are more time-consuming, their advantages over narrow dipstick assessments are:

- Their ability to do justice to the paces at which children's competencies develop – for example, in the TIME Study sample, isiXhosa-speaking children in Grade R counted less fluently than English-speaking children but had superior reasoning skills. This would not have been noticed if we had used a dipstick assessment focused on counting and basic arithmetic.
- Broad-scope assessments help counter the risk of teaching-to-the-test, which may take place at the expense of very important competencies (e.g. decoding at the expense of comprehension, reading at the expense of writing).



CONDITIONS FOR SUCCESSFUL USE IN DIFFERENT CONTEXTS

WELA and MELA are useful tools for **small-scale** diagnostic **studies** as they provide rich insights into the various facets of the child's competency. The WELA tool, with its set of questions that evolve over time, is also useful for **short longitudinal studies**, provided that the following conditions are met:

- Work with a limited number of assessors and enforce consistent assessment protocols across assessors.
- Have a quality control process, possibly with a scoring panel, to harmonise the variable scoring practices. Despite the provision of detailed scoring guidelines, inconsistencies are still likely to creep in, particularly for the scoring of stories.

NGOs have also used the WELA tool for **programme monitoring and evaluation**, assessing participating children at multiple intervals (baseline, midline, endline). This use, especially for large-scale programmes, is likely to require:

- A standardised, rather than customised, administration of the assessment: the same predetermined set of tasks is applied to the full sample regardless of stop rules.
- Some routine score checks to strengthen the scoring consistency
- Data cleaning, in particular to avoid incorrect interpretation of zero scores.

The full series of TIME Learning Briefs and other Wordworks publications are available online under: www.wordworks.org.za/insights

WAY FORWARD FOR WELA AND MELA

Both assessment tools are a work in progress. Areas that could benefit from further development, curating and support, include:

- Developing a Grade 1 extension for MELA
- Versioning the tools into other South African languages
- Validating all assessment versions (functional, linguistic, cultural and metric equivalence)
- Formalising and extending the digital data collection process and data management system
- Psychometric analyses and generating norms and standards
- Formalising a training package for assessors and data users
- Licensing
- Website access

In all these areas, it would be important to strike the correct balance between the quest for quality and the less formal character of WELA and MELA, which is intended to keep lower barriers to access than other tools. The positioning and continued relevance of WELA and MELA within the changing landscape of early literacy and numeracy assessments also needs to be constantly kept in mind.

A partnership of Wordworks and RED INK with a well-capacitated, specialised partner could ensure that the tools are taken to their full potential.