AOSIS

Diagnostic test for number concept development during early childhood

CrossMark
click for updates

Author:
Hanrie S. Bezuidenhout[1] ●

Affiliation:
[1]Centre for Education Practice Research, University of Johannesburg, South Africa

Corresponding author:
Hanrie Bezuidenhout,
devilliers.hanrie@gmail.com

Background: This article presents the Afrikaans translation of an originally German diagnostic test for early number concept development. The process of 'importing' a test to South Africa by considering linguistic-, functional-, cultural and metric equivalence is outlined. A theoretical model describes five levels of young children's hierarchical number concept development which collectively contribute to early mathematical understanding. The five-level hierarchical structure has previously been confirmed by testing the theoretical model in a one-dimensional Rasch analysis in Germany.

Aim: The current study aimed to determine whether the individual items, allocated to test the concepts of each level of the Afrikaans translation of the diagnostic test, confirm the hierarchical structure of the theoretical model.

Setting: A Rasch analysis indicated that the model was fit for the Afrikaans translation. A sample of 165 Afrikaans-speaking grade one children was tested in six Afrikaans medium primary schools in Gauteng.

Methods: Analysis of fit values, person and item reliability and a person–item map was used as part of a Rasch analysis.

Results: The theoretical model of hierarchical number concept development holds for the Afrikaans MARKO-D. Five levels were clearly distinguishable on a Write map and the individual items tested the concepts according to the levels of the theoretical model.

Conclusion: The Afrikaans MARKO-D can now successfully be used to describe young Afrikaans children's number concept development. A five-level theoretical model is a useful tool for teachers using the MARKO-D to assess young children's numerical competence.

Introduction: Background and orientation of the study

Young children's understanding of early number concepts varies significantly in the early school years (Desoete 2015). Results of international tests, such as the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ 2010) and Trends in International Mathematics and Science Study (TIMSS 2015), highlight the idea that many young children in South Africa struggle to develop mathematical competence during their first few years of education. Taylor and Vinjevold (1999) summarise the findings of these assessments:

> At all levels investigated by [The President's Education Initiative], the conceptual knowledge of students is well below than expected at the respective grades. Furthermore, because students are infrequently required to engage with tasks at any but the most elementary cognitive level, the development of higher order skills is stunted. (p. 231)

Previous studies have indicated that there is a relationship between early number concept development and mathematical performance when children progress through formal education (Aunola et al. 2004). Poor results on early number concept assessments are therefore worrying. Young children's number concept understanding should ideally be described at the beginning of formal education to identify children who may possibly encounter mathematical learning difficulties (Mononen & Aunio 2016).

For foundation phase teachers to teach number concepts and identify children who may struggle to develop such concepts, pedagogical content knowledge of the developing child is required (Henning 2013). Knowledge about policy, teaching methods, curriculum and annual national assessments in South Africa (as Henning [2013] found to be the discourse of foundation phase teachers' knowledge) is not sufficient to teach young children. Rather, child cognition and early

number concept development should be 'the primary content of foundation phase teachers' knowledge' (Henning 2013). Coupled with teachers' knowledge, a diagnostic instrument which can describe young children's early number concept development can be utilised in (or prior to) grade one to identify children with possible difficulties.

Unfortunately, Venkat and Spaull (2015) found that most teachers do not have adequate knowledge of mathematics or knowledge of the developing child (Henning 2013). Coupled with the likelihood of a lack of knowledge, most teachers in South Africa do not have access to diagnostic tests to identify children with possible mathematical difficulties (Fritz et al. 2014). One could argue that the Annual National Assessment (ANA[1]) test was designed 'to determine learner performance with regard to the skills and knowledge that they have acquired as a result of teaching and learning experiences in school' (Department of Basic Education 2013:7). However, 'these tests were not psychometrically calibrated to be comparable across time or between grades making them unusable for comparison purposes' (Spaull & Kotze 2015) and although the ANAs were able to indicate the socio-economic learning gaps in South Africa (Van den Berg 2015), the tests were not successful in providing information on learner performance. Alternatively, another curriculum-based test, the VASSI (Vassiliou 2003), which describes children's understanding of concepts determined by the school curriculum in comparison with their classmates, could be used. However, the VASSI only provides a numerical summary of a child's performance in relation to a norm and do not describe which concepts the child has already learnt and which concepts are yet to be developed (Fritz et al. 2014).

In preference to curriculum-based tests, this article presents the MARKO-D (Mathematical and Arithmetic Competence Diagnostic, as translated from German) which is a diagnostic instrument with the ability to describe children's conceptual understanding. I argue that the Afrikaans translation of the MARKO-D, together with the theoretical model which forms the foundation of the test, can support Afrikaans teachers to assess, describe and teach number concepts to young children. The purpose of this article is to answer the research question: Does the theoretical model of hierarchical number concept development hold for the Afrikaans translation of the MARKO-D-test? Three sub-questions guided the researcher: (1) Can the five development levels of the Afrikaans MARKO-D be distinguished on a Wright map?; (2) Do the individual items test the concepts of the levels it was designed to test?; and (3) Is the Afrikaans translation an accurate translation of the German test?

Development, translation and adaption of the Mathematical and Arithmetic Competence Diagnostic

In search for an instrument that can describe young children's number concept development, a research community in Johannesburg encountered the German MARKO-D test

(Ehlert & Fritz 2013) which they decided to translate into English, Afrikaans, isiZulu and Sesotho for use in South Africa. This article reports on the Afrikaans translation of the test. I begin this section with a description of a model that describes levels of early number knowledge on which basis the MARKO-D was developed. I then turn to a discussion of the translation and adaption of the MARKO-D.

Development of the test

The MARKO-D is informed by a comprehensive model of contemporary theory – the Fritz model (Fritz et al. 2012; 2014, Fritz, Ehlert & Balzer 2013). The starting point for the development of such a theoretical model is the assumption that all children are born with a sense for numbers which develop into more sophisticated cognitive structures (Langhorst, Ehlert & Fritz 2012). Although the process of learning is an inductive process, there are some sort of development levels which build hierarchically on previously developed conceptual structures. Each level of development is characterised by the development of specific concepts that build cumulatively on each other (Fritz et al. 2012). The basic idea of the conceptual model is for a complex skill (such as arithmetic skills) to be broken down into smaller components, to arrange the smaller components on a hierarchical continuum and to develop a set of items that are designed to test specific skills of that particular level. During the test design, possible test items were reduced to a number of 55 items[2] which test the five levels of development. With these items, the model has been tested empirically (Fritz et al. 2014). This section of the article contains a description of the five-level hierarchical model of number concept development.

Level one: Counting

When children are born and language and symbolic counting is still absent, mathematic structures are still very primitive and undeveloped (Dehaene 2011). However, research shows that children are endowed with innate core mathematical structures including an object tracking system (OTS) and an approximate number system (ANS[3]) (Carey 2009; Dehaene 2011; Feigenson, Dehaene & Spelke 2004) which later develop into more sophisticated numerical systems of concepts (Carey 2009). As children grow older, number discrimination increases in precision and their understanding of language develops, but at least for the first year children do not understand the meaning of mathematical language such as counting words (Lipton & Spelke 2003). Although the misconception exists, especially among parents, that toddlers understand what the 10 numerals mean as they begin to imitate the numerals, the correct recitement of these words is only an imitation of 10 meaningless words children hear repeatedly (Sarnecka & Carey 2008).

Only when children follow the three how-to-count principles that Gelman and Gallistel (1978) describe, they are able to

2.The MARKO-D went through another round of adaption after the initial Afrikaans translation. The number of test items was reduced to 47.

3.Some researchers such as Dehaene (2011) refer to the approximate number system and others such as Carey (2009) refer to the same system as the analogue magnitude system.

1.Annual National Assessment.

count with understanding. These include (1) the *one-to-one-correspondence* principle which states that children can enumerate a set by assigning one numeral to each item in a set; (2) the *stable-order* principle which states that numerals are used to count and must always be used in the same order; and (3) the *cardinal principle* which states that the numeral provided to the final item in the set describes the number of items in the set. At this point children start to count out a specific number of elements, successfully link each number to individual objects or their fingers (one-to-one match), count small collections of objects where the counting process ends with the last number recalled. This is the first level of the theoretical model which informs the MARKO-D: counting. Table 1 presents examples of MARKO-D items of each level.

Level two: Ordernality

At this point, children do not only know a rhyme for counting in a specific order but also begin to develop an ordinal representation of numbers that are increasing in size owing to their position on the number line (Fritz et al. 2012). Children begin to understand that numbers that occur later on in the count list represent bigger cardinalities and that there is a specific order to the numerals (Sarnecka 2016). To understand the order of the numbers in the count list, children must also understand the successor principle which states that 'one more' than each number refers to the 'next' number (Sarnecka 2016). Now, children can compare numbers and determine which number is smaller or bigger. They can do simple addition and subtraction calculations by using the number line. At this level, children will use different counting strategies to do these simple addition and subtraction calculations (Fritz et al. 2014).

To test children's knowledge of ordernality, Sarnecka and Carey (2008) conducted *the Direction task* where they showed children two plates, each containing six items. One item was removed from one plate and placed with the six items on the other plate. The children were asked: 'Which plate has five?' From this experiment they concluded that some children 'did not say that the set that had gained something should now have a number from later in the list, while the set that lost something should have a number from earlier in the list' (Sarnecka 2016:162). Some children understood the principle of ordernality and others did not understand order or cardinality of numbers.

Level three: Cardinality

For real cardinal understanding, children should understand that each word not only refers to a number of elements in a set but also that these elements form a set of a specific quantity (Sarnecka & Wright 2012). At this level, children fill the placeholder structure of the counting list with meaning – the

conceptual understanding of how counting represents number. True cardinal understanding thus refers to the mental integration of the relevant number name and the number of objects that is counted so far (Fritz et al. 2014). In other words, language symbols and concepts meet in representation (Kozulin 1990). If, for example, eight objects are counted, a number name is assigned to each object and all objects are integrated to determine the total of eight. On the third level of the theoretical model a child understands, for example, the 'eightness' of eight. To assess children's understanding of cardinality, Wynn (1990; 1992) designed the Give-N (or Give-a-Number-task) in which children are asked to present a set of a specific number. For example, a child may be given 15 objects and asked to give 5 of the items to a stuffed animal. Children who were able to present five objects were likely to understand the cardinality of five, whereas children who were unable to produce a set of five perhaps did not yet understand the cardinality of five. In the MARKO-D, children are asked to make sets equal by adding a number of objects. For this task, they must understand the principle of cardinality.

Level four: Part-part-whole

At the next level, children learn that all numbers can be decomposed into smaller units but also form part of larger quantities (Langhorst et al. 2012). The understanding that all numbers are compositions of other numbers allows understanding of operations (Langhorst et al. 2012). Resnick (1983) considers the ability to interpret quantitative relationships through the part-part-whole concept as the most important mathematical acquisition during early childhood. At this level, children begin to determine the relationship between the *part-and-part* and the *whole* which enables them to flexibly do operations. For example, in the MARKO-D children are asked: 'Give me five counters. Three of them must be red!' To solve this problem, children understand that five can be decomposed into three and two. Three of the five counters must be red and the remaining two must be blue.

Level five: Relationality

The fifth level requires a deep understanding of the complex concept of natural numbers (Fritz et al. 2014). At this level, children integrate their understanding of ordinality, cardinality and the relationship between numbers as explained by level four (Fritz et al. 2012). The count list is now understood as a sequence of units, where each number in turn is an independent countable unit. The child realises (according to the modelling of this level) that numbers and counting skills can not only be used to count concrete objects but that numbers also refer to abstract quantities and units. In the MARKO-D, children are asked: 'Give me eight counters. There must be two more red ones than blue ones.' For this task, they must understand the relationship between numbers.

Translation and adaption of the test

When studying child development in South Africa, research often requires the translation of instruments and instructions

TABLE 1: Summary of Rasch analysis.

Variables	Fit- and reliability statistics
MNSQ infit	1.00
Person reliability	0.88
Item reliability	0.97

Note: Input: 165 persons 55 items. Reported: 162 persons 55 items.
MNSQ infit, mean-square infit values indicate model fit.

owing to various national languages. The translation process not only requires linguistic equivalence but also functional-, cultural and metric equivalence (Peña 2007). To translate the German MARKO-D to Afrikaans, an inductive process of many rounds of translation and back translation was followed. Firstly, the original German test was translated to English. I then translated the English test to Afrikaans. A translator who is fluent in Afrikaans, English and German back translated the Afrikaans and English tests to German to determine linguistic accuracy. The original test and back translated German tests were compared to identify and resolve differences. Next, various Afrikaans-speaking foundation phase teachers reviewed the Afrikaans translation to ensure its accuracy. The goal of linguistic equivalence is to ensure that the linguistic meaning is the same for both instruments (Grisay 2003).

Peña (2007) highlights that linguistic equivalence is not sufficient and may introduce bias. Even if the words are translated accurately, potential subtle differences may 'result in different patterns of responses' (Peña 2007:1257). Therefore, functional-, cultural and metric equivalence were considered during the translation of the MARKO-D. Functional equivalence, which is often overlooked, allows assessment of the same construct (Peña 2007). During the translation of the MARKO-D, the focus was to ensure that all items initially designed to measure specific concepts in the German test (as described in the theoretical model), measured the same concepts in the Afrikaans translation. In this regard, functional equivalence is closely related to metric equivalence which refers to comparability of item difficulty between the two languages (Peña 2007). The aim of this study was to determine whether the Afrikaans translation of the MARKO-D was equivalent in item difficulty and whether the concepts of the five levels measured in the German test, measured the same concepts in the Afrikaans test – as described by the theoretical model. In Afrikaans and English, we were able to translate the items in such a way that the item difficulty corresponded so as to explain the principles of equivalence. I will use an isiZulu example as the isiZulu and Sesotho translations were more difficult to translate. For instance, if one would ask: 'What is one more than ten' an English child would have to know the name 'eleven'. For an isiZulu child this would be a redundant question since 'yishumi nanye' (11) translates as 'one more than ten'. The item difficulty would most likely differ between English and isiZulu.

Lastly, cultural aspects had to be considered. Cultural equivalence focuses on the way in which individual children interpret the underlying meaning of items which may affect responses to items and standardised tests (Canino & Guarnaccia 1997; Hendrickson 2003). In Germany, the MARKO-D was designed to include a storyline of animals communicating with each other about numbers. During an individual interview-based assessment, the child became part of the conversation by providing answers for questions posed by some of the characters, by answering questions about the surroundings or by acting as one of the characters. In the German test, the storyline is about squirrels who are preparing for the winter time by collecting nuts and berries. To most South African children, this is an unfamiliar context owing to weather and cultural differences between South Africa and Germany. Therefore, the South African team that translated the MARKO-D for local use adapted the storyline to address cultural differences.

During the adaption, characters and the story line had to be considered. Animals like dogs, cats, elephants and rabbits were considered, but because the test was also translated to isiZulu and Sesotho, all cultural differences had to be taken into account. This process included deciding on characters all children in South Africa could relate to – not dogs or cats that could be aggressive animals for some and pets for others; not cows, pigs or any other animal that could possibly be holy or unholy for some children because of their cultural and religious views, but a neutral, friendly 'African' animal. Finally, meerkats were chosen. The South African story was written in collaboration with the authors of the German story and the pictures were drawn by a local art teacher.

Research methods and design
Sample

To test equivalence between the Afrikaans and original MARKO-D, 165 Afrikaans grade 1 children with Afrikaans as home language were assessed during the beginning of the grade one academic year. Seventy-eight children were tested in one of Johannesburg's Afrikaans elite private schools and 20 in another Afrikaans school with good socio-economic circumstances. These children were tested during February and March. The remaining 67 children were tested by JET research company during April and May. These 67 children all attend one of four Afrikaans schools with Afrikaans as their home language. The children were selected by purposeful intact group selection. Data of three children were incomplete and therefore only data of 162 children were used for analysis.

Test instrument

The MARKO-D consisted of 55 items and 13 sub-sections. Each item tested a specific number concept which is described by one of the levels of the theoretical model of conceptual development (Ricken, Fritz & Balzer 2013). Test administration was conducted in individual interview style with each session taking approximately 40 min. Each interviewer used a booklet which included an exact dialog which the tester strictly followed to ensure unity among testers and pictures of the story of the meerkats which were presented to the children. The responses of the participants were recorded on a recording sheet.

Rasch analysis as analytical tool

To determine whether the Afrikaans translation fits the theoretical model which was used to design the original German MARKO-D, I utilised Rasch modelling (Bond & Fox 2007). This type of probability model aims to accurately determine the possibility of a specific person reacting in a

certain (predicted) way to the various items on a test instrument (in order of difficulty according to the theoretical model). Such a statistical model can be used to determine whether an instrument (such as the Afrikaans MARKO-D) meets the specific requirements of the probability model. If the test instrument fits the model by meeting the requirements of the five levels of conceptual development, the statistical model may be considered appropriate to make meaningful and valid conclusions about, for example, the equivalence of the diagnostic mathematics test for the foundation phase.

Rasch modelling allows one to simultaneously represent the degree of difficulty of items (which measures various levels of a single construct, such as number concept development) and a person's ability (according to which level's items the person was able to answer correctly) on a visual model (Bond & Fox 2007). Figure 1 explains this idea of Rasch modelling (adapted from Balzer 2014).

One of the principles of Rasch modelling is the idea of measuring only one construct (or latent trait) at a time (Bond & Fox 2007). This is referred to as unidimentionality. On a Wright map (also called a person–item map), the vertical line represents the latent trait: number concept development (as illustrated in Figure 1). On the left hand side, the persons' ability is represented. For instance, person A has a low ability and was unable to answer any of the items correctly (all items on the right are represented above person A). Person B was able to answer some of the questions and person C has a high ability and was able to answer all the items correctly (all the items on the right are lower than person C). In other words, the higher the person is represented on the map, the more competent that person is. The difficulty of the items is represented on the right of the vertical line. The items at the bottom, such as item 1, are the easy items because most persons were able to answer them correctly. Item 2 is more difficult and item 3 is a very difficult item, because only person C was able to answer that specific item correctly.
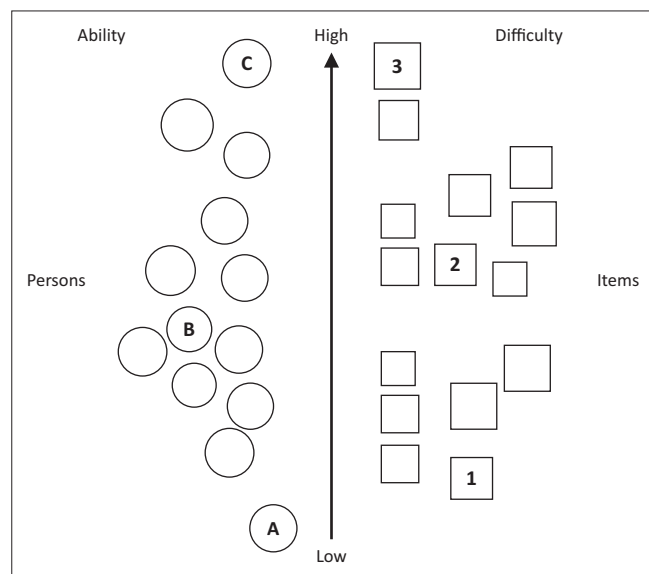
Source: Adapted from Balzer's explanation of Rasch modelling (Balzer 2014)
FIGURE 1: Item difficulty and person ability.

Thus, difficult items (high on the continuum) are less likely to be answered by all persons and easier items (lower on the continuum) are more likely to be answered correctly.

A single construct, such as number concept development, can be broken up into smaller concepts (described by the theoretical model) and can be measured according to levels of competence. In the example of Figure 1, three distinct levels can be distinguished. For the sake of explaining the idea of levels, the three levels in Figure 1 are much more distinguishable than in reality. One could easily say that the items clustered around item 1 can be grouped into one level, the items clustered around item 2 can be grouped into another and the items around item 3 can be seen as a third level of difficulty. In Figure 2, a person–item map of this study, it is clear that number concept development is a developmental hierarchy with less distinguishable levels.

Another principle of Rasch analysis is that each item must contribute to the construct being measured in a meaningful way (Bond & Fox 2007). In Rasch analysis, fit statistics allows us to determine how well each item fits within the single construct. Bond and Fox (2007:35) explain that fit statistics 'help to determine whether the item estimations may be held as meaningful quantitative summaries of the observations'. They (Bond & Fox 2007) compare fit statistics to confidence levels. The closer the fit statistics (infit values) are to 1, the better the item fits the model.

The third principle of Rasch analysis is person and item reliability. Person reliability indicates the replicability of the order of persons one could expect if the same sample of persons were given a similar set of items (Wright & Masters 1982). Item reliability indicates the replicability of the placements of items on the continuum if another set of persons were given the same items (Bond & Fox 2007). A reliability value close to 1 is a high reliability value.

Ethical considerations

Each school's management team and the Gauteng Department of Education gave permission for the research to be conducted in public schools. Participants' parents also gave consent for their children's participation in the research. The identity of the children and the results were handled confidentially.

Ethical number: 2012-04

Results

Excel was used to create a raw data set in which '1' indicated a correct response and '0' an incorrect response (dichotomous data). The data were cleaned by checking whether all the data have been captured correctly and by making sure there was no missing data. Thereafter, I imported the data to WINSTEPS 3.81.0 to perform a Rasch analysis (Bond & Fox 2007) in order to produce two tables with fit values (Bond & Fox 2007) and a Wright map (Bond & Fox 2007). The purpose of the first table is to indicate whether the Rasch model was a suitable model for data analysis.
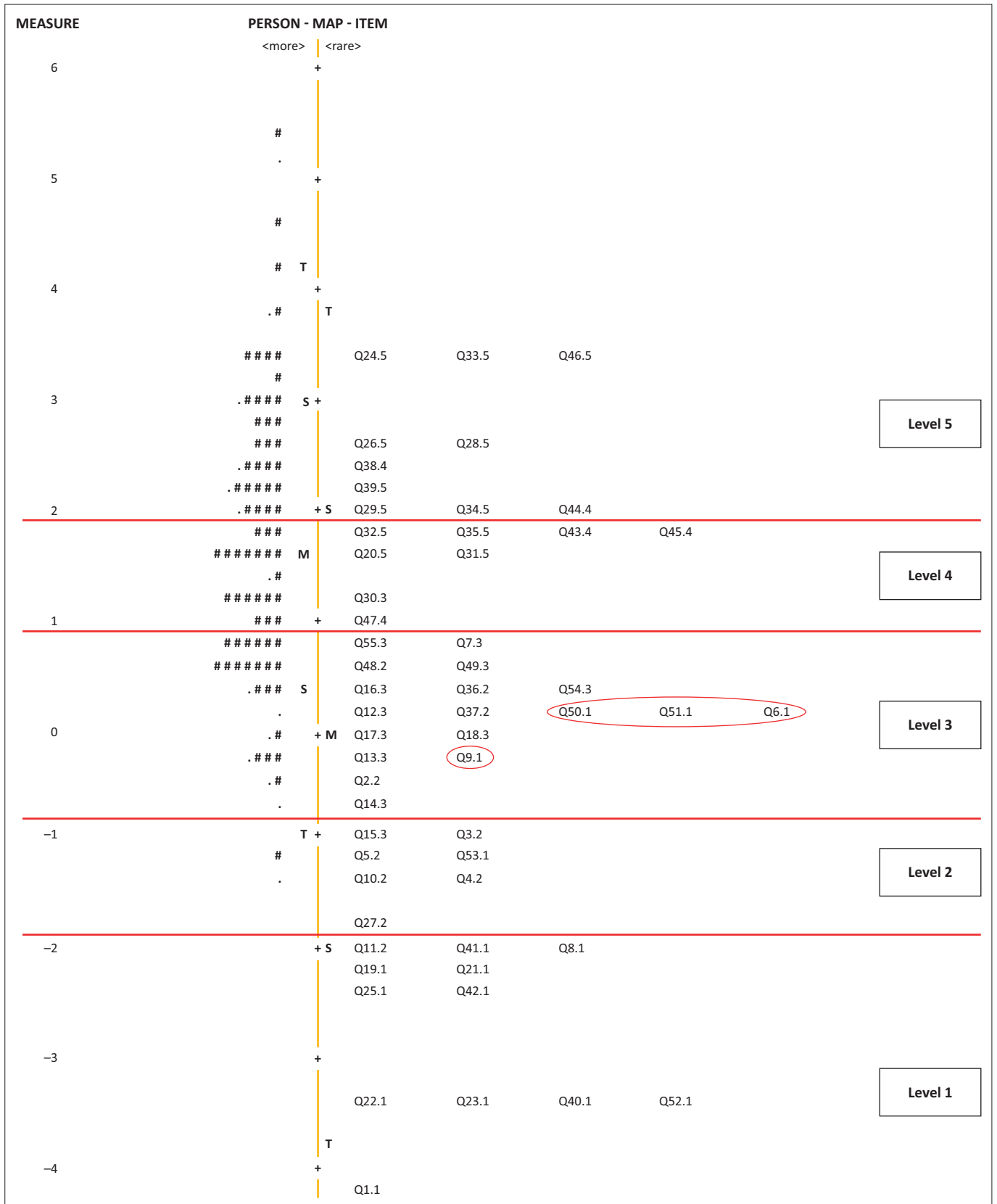
FIGURE 2: Afrikaans Mathematical and Arithmetic Competence Diagnostic's spread of children and items on a Wright map.

Table 1 indicates that the *person reliability* is 0.88 and the *item reliability* is 0.97. Standardised infit- or oufit values, or mean square values (MNSQ) close to 1, indicate a good model fit. The infit MNSQ for this study was 1. Therefore, I concluded that the Rasch model was suitable to analyse the data set at hand. The second table produced by the Rasch analysis had the purpose to indicate the fit values for each item in order to determine whether each item did indeed test what it was

supposed to test. In Table 1, 'higher MNSQ values indicate too low selectivity; lower MNSQ values indicate too high selectivity, and thus the presence of redundant items in the test' (Fritz et al. 2013:55). Wright and Stone (1979) suggest 1 ± 0.5 as an acceptable range for 'infit' MNSQ values. Wright and Linacre (1994) indicate that 1 ± 0.2 are valid limits for acceptable 'infit' values and 1 ± 0.3 are less strict but still sufficient limits.

Table 2 indicates that 52 items' fit values were between 0.8 and 1.2, and two items' fit values were between 0.7 and 1.3. Item 38 had a value of 1.24 and item 13 a value of 0.7. Item 36 had a value of 1.34, which was just outside the values Wright and Linacre (1994) suggested but still within the suggested values of Wright and Stone (1979). Item 36 was examined as this item's value was just above the suggested fit values. The wording of the particular item was changed for the next round of testing, and after the second round of testing all of the values were within the suggested limits.

During the standardisation of the test in Germany, Lars Balzer (one of the test designers) has developed Table 3 to indicate which concepts, according to the development levels, each item is supposed to test. The levels were statistically determined during the standardisation process in Germany.

I agree with the classification of most of the levels of the items. However, I would argue from a theoretical point of

TABLE 2: Fit statistics for the individual items.

Item	MNSQ infit	Item	MNSQ infit	Item	MNSQ infit
41	0.99	18	1.06	34	0.90
1	1.04	19	1.07	51	0.96
21	1.10	49	0.98	4	0.94
53	1.01	26	0.89	20	0.94
36	1.34	2	1.04	52	0.94
42	1.05	23	1.05	5	0.93
37	1.14	39	1.05	12	0.93
6	1.14	22	1.04	17	0.93
40	.97	8	1.03	55	0.93
54	1.17	11	1.01	14	0.91
9	1.19	27	1.01	48	0.91
38	1.24	47	1.01	33	0.90
25	1.08	3	0.97	15	0.89
16	1.12	29	1	44	0.87
46	1.17	32	1	28	0.85
30	1.11	10	0.98	45	0.84
50	0.97	15	0.98	43	0.86
24	1.02	31	0.96	13	0.70

Note: Input: 165 persons 55 items. Reported: 162 persons 55 items.
MNSQ infit, mean-square infit values indicate model fit.

TABLE 3: Mathematical and Arithmetic Competence Diagnostic items according to development levels.

Development level	Item number	Items
1	1	Count as far as you can.
2	2	What comes just before 5?
2	3	What comes just after 5?
2	4	What comes between 2 and 4?
2	5	What comes between 5 and 7?
1	6	Count the stones. How many are there?

TABLE 3 (Continues...): Mathematical and Arithmetic Competence Diagnostic items according to development levels.

Development level	Item number	Items
3	7	How many stones are there?
1	8	Count the stones.
1	9	How many are there?
2	10	My sister has two apples. Grandmother gives her two more apples. How many apples does my sister have now?
2	11	Can you show me how to do this with the counters?
3	12	Beni says: 'I had five eggs and my brother ate three. How many do I have now?'
3	13	Can you show me how to do this with the counters?
3	14	Beni says: 'I put three nuts in a hole and my sister puts three more nuts in the hole. How many nuts are there in the hole?'
3	15	Can you show me how to do this with the counters?
3	16	How many counters have to go in the empty block? 1 [---]-3-4-5
3	17	How many counters have to go in the empty block? [---]-3-4-5
3	18	How many counters have to go in the empty block? 4-[---]-6-7
1	19	Which row has more? 6-3
5	20	How many more are there? 6-3
1	21	Which row has less? 6-4
1	22	How many less are there? 6-4
1	23	Which row has less? 9-4
5	24	How many less are there? 9-4
1	25	Which row has less? 6-5
5	26	How many less are there? 6-5
2	27	Which row has more? 6-7
5	28	How many more are there? 6-7
5	29	What is one smaller than five?
3	30	What is one bigger than seven?
5	31	What is one smaller than nine?
5	32	What is two bigger than three?
5	33	What is three smaller than seven?
5	34	What is two smaller than five?
5	35	What is two bigger than four?
2	36	Here you can see four stars and three stars are hiding behind the cloud. How many stars are there altogether?
2	37	Here you can see five stars and three stars are hiding behind the cloud. How many stars are there altogether?
4	38	If I count 1-3-5, how does it go on?
5	39	We can also count backwards like this: 10-8-6. How does it go on?
1	40	Give me four counters.
1	41	Give me six counters.
1	42	Give me five counters.
4	43	Give me five counters. Three of them must be red.
4	44	Give me nine counters. Four of them must be blue.
4	45	Give me six counters. There must be more blue ones than red ones.
5	46	Give me eight counters. There must be two more red ones than blue ones.
4	47	I want ten nuts. I already have four. How many must I still get to have 10?
2	48	Lona has eight nuts. How many must she take away so that there are only six?
3	49	I have some counters under my hand. I take away three. Now I have five. How many counters were there under my hand before?
1	50	Divide 10 apples between the two rabbits so that each rabbit has the same number of apples.
1	51	Divide eight apples between the two rabbits so that each rabbit has the same number of apples.
1	52	4-0: Put down here as many counters as you can see here.
1	53	6-0: Put down here as many counters as you can see here.
3	54	8-4: Put down here as many counters as you can see here.
3	55	7-2: Put down here as many counters as you can see here.

view that items 7, 36, 37, 48, 49, 50 and 51 test concepts from different conceptual levels, as indicated in Table 3. For example, item 7 (How many stones are there?) is said to test concepts of level three (cardinality), but the ability to count a set of objects could rather be classified as a level one concept. Similarly, items 36 and 37 (Here you can see four or five stars and three are hiding behind the cloud. How many stars are there altogether?) are said to test concepts of level two, but I would argue that these are examples of questions of level four (part-part-whole). I would also classify items 48 and 49 as items that test concepts of level four rather than levels two and three and items 50 and 51 as items from level three rather than level one.

To look at the spread of the items statistically by using the current sample, I used WINSTEPS 3.81.0 to produce a Wright map (also called a person-item map). Figure 2 is a representation of both persons (#) and items (Q24.5). The vertical line in the middle represents the unit of analysis which was grade 1 children's performance on a diagnostic numeracy test. On the left hand side, the participating children are represented by a #. On the right hand side are representations of the items. Each item indicates the number of the item on the test (Q24.5), as well as the level of conceptual understanding the item was statistically supposed to test (Q24.5), as shown in Table 3. On the Wright map, it is clear that this instrument was too easy for the sample, as the distribution of the children on the left is located higher on the spectrum than the distribution of the items on the right.

From a comparison between the Wright map (Figure 2) and Table 3, it is clear that the positioning of certain items on the map differs from the level it is designed to test. This means that items like items 50 and 51 statistically (according to the standardisation process in Germany) should test concepts of level one (count numbers). However, according to the Wright map (Figure 2), which has been produced by WINSTEPS 3.81.0 (Linacre 2012) during a Rasch analysis, these items test concepts in the third level of development which tests the principle of cardinality. It may be that the strategy (procedure) that children in Grade R learn to solve these problems brings about this difference.

There are four items (items 9, 50, 51 and 6) of which the development levels on the Wright map and Table 3 differ significantly. This means that the distance between the level that the item is supposed to test (according to Table 3) and the level which is shown on the Wright map is more than one level apart. There are several possible reasons why these (easy) items for the group of pupils tested during this study were more difficult than it was supposed to be. During the administration of the test, I noticed that many children forget in item 6, where they started to count. As a result, they count one or two stones twice or exclude one or two stones when they count Lona's stones. If the dots (or stones) are placed in a different arrangement, it is likely that children will be able to count it more easily. However, the test is

designed to test whether children can count small amounts in all formats (see Piaget 1965) and therefore this item was not changed during the adjustment of the test.

Item 9 tests the idea that the amount of stones does not change when you start counting elsewhere, but it seemed that children did not understand what the question meant when the test administrator asked: 'How many stones are there if I start counting here?' while pointing to a different starting point to count the number of items. Item 9 was eliminated from the test during adjustment of the test. For items 50 and 51, I would argue from a theoretical position that halving numbers does not belong in the first level of development, but at the third level (cardinality).

Regarding the other 13 items (11, 20, 37, 15, 53, 2, 36, 48, 30, 32, 35, 44 and 38) which has a difference between the levels of development on the Wright map and Table 3, I argue as follows: because the theoretical model used in this study as an underlying model is a *development* model, it is acceptable if an item on the Wright map is one level higher or one lower than indicated in the table. For example, most of the items on the Wright map in level one are items that are supposed to test the concepts of level one, except item 11 which is an example of an item that is supposed to be in level two. Because it is a *developmental* model without strict boundaries, I argue that it is acceptable if item 11 is placed high in the first level.

Conclusion

The objectives of this study was to determine whether the five development levels could be distinguished on a Wright map, to find out whether the individual items tested the concepts of the level it was designed to and to determine whether the Afrikaans translation was an accurate translation of the German test. The fit tables and Wright map presented in this article indicate that most items did indeed test the concepts we hypothesised it would test within the five distinguishable levels described by the theoretical model. Also, through a process of back translation with linguistic-, functional-, cultural and metric considerations, the research team ensured an accurate translation of each item. Therefore, I conclude that the theoretical model of the hierarchical number concept development holds for the Afrikaans translation of the MARKO-D-test.

This article has implications for research, practice and policy. Through this study, an Afrikaans assessment instrument is now available to assess and describe young children's numerical skills. The theory presented here could be used during teacher training and teacher development programmes in order to improve foundation phase teachers' knowledge about early number concept development. Green (2017) emphasises that South Africa needs 'theoretical teachers' whose practice is informed by theory which describes the developing child. From my experience as a foundation phase and mathematics teacher, I know that not many teachers utilise

the theory of hierarchical concept development to inform their teaching because they do not know such theory. I claim that teachers and curriculum designers can use the information in the field of developmental psychology to enrich their own knowledge about learning and adapt the way they teach (their pedagogy). According to Henning (2013), unless child cognition is the primary disciplinary content of the knowledge of foundation phase teachers, teachers might 'couch their pedagogical knowledge in teaching methods and materials more than in knowledge of conceptual development of learners' (Henning 2013:139). All role players – teachers, teacher trainers, curriculum designers and researchers – should thus include theory of child cognition in their own content knowledge.

Because a diagnostic test is never meant to be used in isolation (Mononen & Aunio 2016), but rather in collaboration with an intervention programme, future research should focus on the development and implementation of an intervention programme with the goal to support children who do not understand the numerical concepts appropriate for their age (according to the MARKO-D results). Such an intervention programme could also use the Fritz model as a basis to develop number concepts hierarchically during intervention. Future research should also focus on the development of instruments that could be used together with the MARKO-D to assess young children's cognitive skills, which could contribute to early number concept development, such as language and executive functions (Purpura et al. 2017). Should researchers consider to 'import' existing tests, this study provides a framework for considerations during test adaption and translation. Although preliminary analysis has indicated that the concepts of the MARKO-D correspond with the concepts in the South African curriculum, further investigation could yield a better insight into the South African curriculum that incorporates the concepts of the MARKO-D.

One of the shortcomings of the study is the question whether a 40 min 'individual interview style' test is suitable for universal implementation in South Africa. In this regard, the original test designers from Germany are currently considering the option of a short version of the MARKO-D to 'screen' children before conducting lengthy 40-min interviews with each child.

Acknowledgements

Competing interests

References

Aunola, K., Leskinen, E., Lerkkanen, M.-K. & Nurmi, J.-E., 2004, 'Developmental dynamics of math performance from preschool to Grade 2', *Journal of Educational Psychology* 96(4), 699–713. https://doi.org/10.1037/0022-0663.96.4.699

Balzer, L., 2014, 'Getting to know Rasch modelling', article presented at Research Methodology Winter School by Centre for Education Practice Research (CEPR), University of Johannesburg, Johannesburg, 23–27th June.

Bond, T.G. & Fox, C.M., 2013, *Applying the Rasch model: Fundamental measurement in the human sciences,* 2nd edn., Erlbaum, Mahwah, NJ.

Canino, G. & Guarnaccia, P., 1997, 'Methodological challenges in the assessment of Hispanic children and adolescents', *Applied Development Sciences* 1, 124–134.

Carey, S., 2009, *The origin of concepts*, Oxford University Press, Oxford.

Dehaene, S., 2011, *The number sense (How the mind creates mathematics),* Oxford University Press, Oxford.

Department of Basic Education, 2013, *Annual national assessment: 2013 Diagnostic report and 2014 framework for improvement,* viewed 28 December 2014, from http://www.education.gov.za/Curriculum/AnnualNationalAssessment/tabid/424/Default.aspx

Desoete, A., 2015, 'Predictive indicators for mathematical learning disabilities/dyscalculia in kindergarten children', in S. Chinn (ed.), *The Routledge international handbook of dyscalculia and mathematical learning difficulties,* pp. 90–100, Routledge, New York.

Ehlert, A. & Fritz, A., 2013, 'Evaluation of maths training programme for children with learning difficulties' *South African Journal of Childhood Education* 3(1), 101–117.

Feigenson, L., Dehaene, S. & Spelke, E., 2004, 'Core systems of number', *Trends in Cognitive Science* 8, 307–314.

Fritz, A., Balzer, L., Herholdt, R., Ragpot, L. & Ehlert, A., 2014, 'A mathematics competence test for Grade 1 children migrates from Germany to South Africa', *South African Journal of Childhood Education* 4(2), 114–133.

Fritz, A., Ehlert, A. & Balzer, L., 2013, 'Development of mathematical concepts as basis for an elaborated mathematical understanding', *South African Journal of Childhood Education* 3(1), 21–37.

Gelman, R. & Gallistel, C.R., 1978, *The child's concept of number*, Harvard University Press, Cambridge.

Green, W., 2017, 'Opening comments by Dr Whitfiel Green, Chief Director for Teaching and Learning Development, DHET' presented at Seminar of Education in Finland: Best practices, The University of Johannesburg, Soweto, 31st May.

Grisay, A., 2003, 'Translation procedures in OECD/PISA 2000 international assessment', *Language Testing* 20, 225–240.

Hendrickson, S.G., 2003, 'Beyond translation… Cultural fit', *Western Journal of Nursing Research* 25, 593–608.

Henning, E., 2012, 'Learning concepts, language, and literacy in hybrid linguistic codes: The multilingual maze of urban grade 1 classrooms in South Africa', *Perspectives in Education* 30(4), 70–79.

Henning, E., 2013, 'Teachers' understanding of mathematical cognition in childhood: Towards a shift in pedagogical content knowledge?', *Perspectives in Education* 31(3), 139–154.

Kozulin, A., 1990, *Vygotsky's psychology: A biography of ideas*, Harvard, Cambridge, MA.

Langhorst, P., Ehlert, A. & Fritz, A., 2012, 'Non-numerical and numerical understanding of the part-whole concept of children aged 4 to 8 in word problems', *Journal für Mathematik-Didaktik* 33(2), 233–262. https://doi.org/10.1007/s13138-012-0039-5

Linacre, J.M., 2012, WINSTEPS Rasch measurement computer program, Winsteps.com, Chicago, IL.

Lipton, J.S. & Spelke, E., 2003, 'Origins of number sense: Large-number discrimination in human infants', *Psychological Science* 14(5), 396–401. https://doi.org/10.1111/1467-9280.01453

Mononen, R. & Aunio, P., 2016, 'Counting skills intervention for low-performing first graders', *South African Journal of Childhood Education* 6(1), 1–9.

Peña, D.E., 2007, 'Lost in translation: Methodological considerations in cross-cultural research', *Child Development* 78(4), 1255–1264.

Piaget, J., 1965, *The child's conception of number*, Norton, New York.

Purpura, D.J., Day, E., Napoli, A.R. & Hart, S.A., 2017, 'Identifying domain-general and domain-specific predictors of low mathematics performance: A classification and regression tree analysis', *Journal of Numerical Cognition* 3(2), 365–399.

Resnick, L.B., 1983, 'A developmental theory of number understanding', in H. Ginsburg (ed.), *The development of mathematics thinking*, pp. 109–151, Academic Press, New York.

Ricken, G., Fritz, A. & Balzer, L., 2013, *MARKO-D – Mathematics und Rechnen – Test zur Erfassung von Konzepten im Vorschulalter* [MARKO-D: Mathematics and arithmetic – Test for assessing concepts in pre-school age], Göttingen, Hogrefe.

SACMEQ, 2010, *The SACMEQ Project in South Africa: A study of the conditions of schooling and the quality of education,* viewed 17 November 2014, from http://www.sacmeq.org/reports

Sarnecka, B.W., 2016, 'How numbers are like the earth', in D. Barner & S. Baron (eds.), *Core knowledge and conceptual change,* pp. 151–170, Oxford University Press, Oxford.

Sarnecka, B.W. & Carey, S., 2008, 'How counting represents number: What children must learn and when they learn it', *Cognition* 108, 662–674.

Sarnecka, B.W. & Wright, C.E., 2012, 'The idea of an exact number: Children's understanding of cardinality and equinumerosity', *Cognitive Science* 37(8), 1493–1506.

Spaull, N. & Kotze, J., 2015, 'Starting behind and staying behind in South Africa: The case of insurmountable deficits in mathematics', *International Journal of Educational Development* 41(1), 13–24.

Taylor, N. & Vinjevold, P., 1999, *Getting learning right: Report of the President's Education Initiative Research Project*, Joint Education Trust, Johannesburg.

Trends in International Mathematics and Science Study (TIMSS), 2015, *TIMSS international database*, viewed 16 April 2016, from https://timssandpirls.bc.edu/timss2015/international-database/index.html

Van den Berg, S., 2015, 'What the annual national assessments can tell us about learning deficits over the education system and the school career', *South African Journal of Childhood Education* 5(2), 28–43.

Vassiliou, C.P., 2003, *VASSI mathematics proficiency test foundation phase*, MindMuzik, Pretoria.

Venkat, H. & Spaull, N., 2015, 'What do we know about primary teachers' mathematical content knowledge in South Africa? An analysis of SACMEQ 2007', *International Journal of Educational Development* 41, 121–130.

Wright, B.D. & Linacre, J.M., 1994, 'Reasonable mean-square fit values', *Rasch Measurement Transactions* 8, 370.

Wright, B.D. & Masters, G.N., 1982, *Rating scale analysis*, MESA Press, Chicago, IL.

Wright, B.D. & Stone, M.H., 1979, *Best test design,* MESA Press, Chicago, IL.

Wynn, K., 1990, 'Children's understanding of counting', *Cognition* 36, 155–193.

Wynn, K., 1992, 'Children's acquisition of number words and the counting system', *Cognitive Psychology* 24, 220–251.